DOCUMENT RESUME

ED 103 468 TM 004 313

AUTHOR Russell, Nolan F.
TITLE Getting Inside the EQA Inventory: Grade 8.
INSTITUTION Pennsylvania State Dept. of Education, Harrisburg. Bureau of Planning and Evaluation.
PUB DATE 74
NOTE 69p.

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS Adjustment (to Environment); Basic Skills; Confidentiality; Creativity; *Educational Assessment; *Educational Objectives; Ethical Values; Grade 8; Health; Junior High Schools; *Measurement Techniques; Scoring; Self Esteem; *State Programs; State Surveys; Student Attitudes; *Testing; Test Reliability; Test Validity
IDENTIFIERS *Educational Quality Assessment; Pennsylvania

ABSTRACT

Although it is difficult both to determine educational goals and to measure goal attainment, the Educational Quality Assessment Inventory (EQAI) has ambitiously tackled both problems. During a five-year period 45 separate instruments containing over 2000 items were constructed, revised, rejected, and accepted in preparation for the present inventory. This inventory has scales which independently measure different facets of educational goals including basic skill achievement, social and health habits, feelings toward self and others, value placed on learning and human accomplishment, interest in creative activities, and methods of coping with freedom. The battery of tests was administered to 253,226 students attending 240 Commonwealth intermediate schools. Though individual names were erased prior to test scoring information necessary to identify general student groups was obtained through questions of sex, ability level, and father occupation. Answers were scored by both norm-referenced and criterion-referenced methods. Selection of these complementary scoring methods enhance the concept that result accountability goes beyond the school. Recognition of many experiences shaping the educational progress of an individual, however, does not detract from efforts to restructure school programs in hopes of goal attainment. (BJG)

Pennsylvania Department of Education 1974

Grade 8

Getting Inside the EQA Inventory

Grade 8

Getting Inside the EQA Inventory

Prepared by
Nolan F. Russell
Educational Research Associate
Division of Educational Quality Assessment
Bureau of Planning and Evaluation
Pennsylvania Department of Education
1974

Commonwealth of Pennsylvania
Milton J. Shapp, *Governor*

Department of Education
John C. Pittenger, *Secretary*

Office of Basic Education
Donald M. Carroll, Jr., *Commissioner*
Harry K. Gerlach, *Deputy Commissioner*

Bureau of Planning and Evaluation
John L. Kennedy, *Director*

Division of Educational Quality Assessment
Thomas E. Kendig, *Chief*

Pennsylvania Department of Education
Box 911
Harrisburg, Pa. 17126

TABLE OF CONTENTS

GETTING ACQUAINTED

The Intermediate Form of the Educational Quality Assessment Inventory (EQAI) for grade 8 is an assessment device designed to give Commonwealth educators meaningful, reliable, and accurate information about their students' development in each of the 10 state adopted quality education goals. The EQAI can be characterized as a collection of highly structured, paper-and-pencil measurement scales. These scales represent an attempt to appraise various aspects of cognitive and human interaction skills together with those attitudes, values, and beliefs thought important in helping our young people adjust to the demands of today's society and tomorrow's world.

Is this a complex task? Yes. Can it be done in such a way as to provide reasonably accurate information? We believe so. However, big words and fancy phrases don't get the job done.

To insure that the scales included in the EQAI could provide relevant information, the Department of Education maintained the stance of testing the tests before using them to test people. After completing the tasks of operationally defining each goal area and developing measurement rationales consistent with these definitions, the department went to the field.

During a five-year period (1969-1974) approximately 45 separate instruments containing over 2,000 items were constructed, revised, rejected or accepted on the basis of countless pilot-tests in over 175 Pennsylvania school districts.

This booklet describes in detail those scales that survived the logical and statistical checks and remain in the final form of the EQAI. Aspects of the descriptions are necessarily statistical and are couched in the language of the test and measurement field. Because the information contained in the publication is equally relevant to school personnel and research scientists, we have made a concerted effort to include explanations of the logic behind the various analytical methods used to determine the adequacy and the efficiency of the tests.

QUESTIONS AND ANSWERS

ABOUT EQA INVENTORY

Who takes the test battery?

The Intermediate Form of the EQA Inventory is designed specifically for 8th grade students in Commonwealth public schools who have a reading comprehension equal to or greater than the average 5th grader. Research indicates that the tests are adequate for students with less reading skill (to grade 3.5) if the tests are read to them. The tests are not recommended for special education students.

How long is the inventory?

Eleven separate tests containing a total of 422 items make up the bulk of the battery. Twelve additional questions obtain student background information including occupation of father or legal guardian, education of mother, sex, size and type of community, stability of residence, etc. Finally 13 questions measure student perception of parental attitude toward the school and home climate. The total battery is typically completed in four sessions of one hour each.

What kinds of tests are in the inventory?

Three multiple-choice instruments tap cognitive skill, achievement and awareness in the areas of verbal analogies, mathematics and vocations respectively. The math and verbal tests are timed. The remaining eight are self-report attitude and interest scales asking students to respond to statements on various continuums such as agree-disagree, true of me-not true of me, yes-no, etc.

What do the tests cover?

Each scale is designed to measure some facet of one state quality education goal. Along with basic skills the various instruments examine 1) social and health habits, 2) feelings toward self and others, 3) value placed on learning and human accomplishments, 4) interest in creative activities and 5) methods of coping with frustration.

Do the tests completely cover each goal?

No. The goals are very broad statements organizing many related concepts under one umbrella. An inventory of 5,000 items probably could not measure the goals in their entirety. Strong efforts have been made to sample some of the most salient facets of each goal. Section One in this book describes in detail which aspects of the goals are measured.

What types of scoring procedures are used?

For the achievement tests simply the number of correct answers is counted. The attitude scales are scored in two ways. First, each set of response options is given a weighting consistent with its corresponding item's direction (i.e. I like school; strongly agree = 3, agree = 2, disagree = 1, and strongly disagree = 0). Item scores are then summed to form a composite score. This is called norm-referenced scoring. The second procedure classifies options into categories of favorable and nonfavorable. In the above example the *strongly agree* and *agree* choices are given a score of one while *disagree* and *strongly disagree* are given a score of zero. Students choosing favorable responses on a simple majority of scale items meet the criterion of minimum positive attitude. This technique is called criterion-referenced scoring.

Can the tests be used to pinpoint specific student-body strengths and weaknesses?

Yes. With the exception of the basic skills instruments, all questionnaires are broken into smaller components called subscales. The inventory's 26 subscales give more specific information than can be provided by the composite scores alone. For example, the Goal V–Citizenship instrument–offers additional scores in the areas of 1) concern for the welfare and dignity of others, 2) respect for law and authority and 3) personal responsibility and integrity. Section One contains descriptions of all subscales.

Are the tests reliable?

Extensive investigation concerning the consistency of student responses within each scale (internal consistency) and the stability of student responses to the scales over time (test-retest reliability) have been conducted by Division of Educational Quality Assessment personnel. All total scales demonstrate high internal consistency reliability and adequate stability. Some of the shorter subscales, however, demonstrate weak internal consistency reliability. Reliability statistics for all subscales and total instruments are presented in Section Two.

Do students fake their answers?

All self-report questionnaires are susceptible to this sort of response bias. During field trials correlations were computed between test scores and a special instrument called the social desirability scale which is designed to pick up the tendency to make oneself look *good*. Where large correlations were found, the tests were revised or dropped. Correlations between *lie* and total instrument scores are presented in Section Two.

Are the tests valid?

Correspondence between ratings made by teachers and the student scores has been demonstrated for six of the attitude scales. Results for a group of studies conducted by the Division of Research coupled with outcomes form a factor-analysis lend further validity support. Findings relating to test validity are presented in Section Three.

How much does testing climate affect final outcomes on the tests?

A 1971 study involving 91 schools showed that the emotional climate (student eagerness, concentration and carefulness) became poorer as the testing session progressed. Correlations between emotional climate and instrument scores, although slightly positive, were not statistically significant. Also schools experiencing adverse testing conditions in terms of settings, distractions, etc. were not found to score lower than schools with no testing distractions on any of the scales.

Are individual student profiles provided?

No. The unit of analysis of all data received from the Educational Quality Assessment Program is the school. No individual student profiles are given. In fact, student names are removed from the answer booklets before being scored as a means of insuring confidentiality of student answers.

Do the tests identify target groups for program focus?

Yes. Even though individual records are unavailable, it is possible to organize data to help identify general student groups having difficulty in a goal area. This is done by summarizing data for various subgroups of students formed from selected student characteristics. The three student characteristics in these analyses are ability level, sex and father's occupation. Section Four shows the proportion of students in each of 18 subgroups who demonstrate positive attitudes on all goal instruments except basic skills and vocational awareness.

Is the EQA Inventory the only source of information for the Educational Quality Assessment Program?

No. In addition to the student questionnaires, there is a survey for teachers and another for school administrators. The results of these surveys are combined to generate a report for each participating school. For a complete description of the contents of these surveys refer to *Manual for Interpreting School Reports.*

What kinds of information does the inventory provide?

Information includes 1) student-body standing on each composite goal test relative to a statewide reference group, 2) student-body standing relative to groups similar in home and school environments, and 3) proportion of student-body who demonstrate minimum positive attitudes.

Are teachers held accountable for poor test scores?

No. The Educational Quality Assessment Program uses three separate assessment inventories to examine student goal achievement at grades 5, 8 and 11. Students at other grade levels do not take the tests. Test outcomes are not solely a result of what teachers at those three levels are or are not doing. Student attitudes and achievements are a complex product of the total home, school and community experience. Accountability only comes into play in terms of taking quality assessment results into consideration when trying to meet the needs of students.

Does the Department of Education offer any help in identifying and implementing curricular strategies that might increase student goal achievement?

Curriculum specialists are investigating new curricular approaches and related literature in the hope of offering interested Pennsylvania schools help to meet student goal needs. As these materials become finalized they are being made available to school districts.

Is there any indication that schools can improve student attitudes by implementing programs?

Yes. Several programs developed by school districts have already yielded measurable improvements on the EQA Inventory scales. The most recent example involves a large district in western Pennsylvania that under an ESEA Title III grant implemented curricular changes which resulted in an increase in their student-body's interest in learning. Specific information about this project is available at the offices of the Division of Educational Quality Assessment upon request.

What information is contained in this booklet?

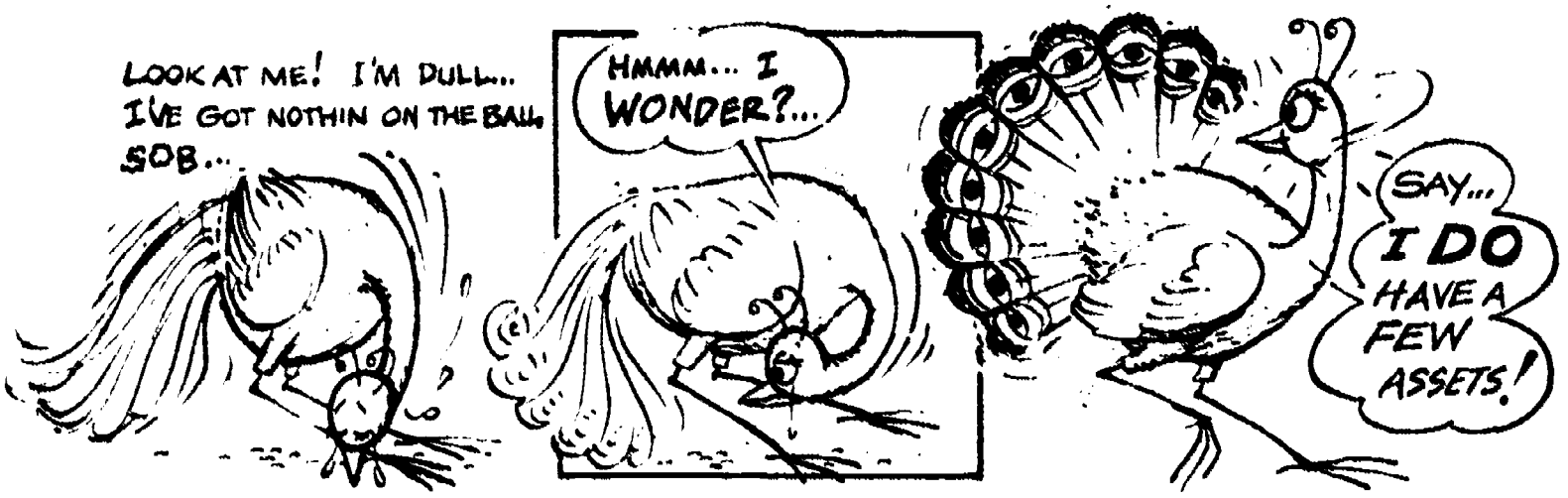
Section One discusses the 10 quality education goals and the measurement devices associated with each. Included in this section are goal and measurement rationales, scale and subscale descriptions, and specifications for scoring. Section Two describes the safeguards used to produce tests of high quality and describes how the EQAI tests fare on these checks. The third section surveys the results of validity studies including teacher ratings, factor analysis and independently conducted studies. Section Four identifies potential student target groups for program focus.

Are there any additional statistical summaries on the tests which are not contained in this booklet?

Yes. This booklet highlights only the major empirical data that are available on the EQA Inventory. Additional materials include item frequency distributions, per cent favorable responses to each item, item-to-total correlations, a complete factor analysis with orthogonal rotations of 2 through 10 factors and various other descriptive statistics including skewness, kurtosis, means, standard deviations and standard errors of measurement for each sub and total scale. These are in computer-printout form and may be seen at the offices of the Division of Educational Quality Assessment in Harrisburg.

SECTION ONE

Goal and Measurement Rationale



GOAL I Self-Esteem

Quality education should help every child acquire the greatest possible understanding of himself or herself and appreciation of his or her worthiness as a member of society.

GOAL RATIONALE

It is widely held that self-understanding is significantly associated with personal satisfaction and effective functioning. How students view their adequacies and inadequacies, their values and desires, can strongly influence their performance in school.

No matter what the level and pattern of students' talents, the school experience should strengthen, not damage, their self-esteem. School should operate so that children of all talent levels can appreciate their worth as persons in a society that claims to be equally concerned for all its members.

MEASUREMENT RATIONALE

Self-esteem is a personal judgment of worthiness. It is a subjective experience which the individual conveys to others verbally or by other behavior. Most theories acknowledge that our self-image and feelings of worthiness are determined largely by how well we can live up to our own aspirations and meet expectations of others.

Aspirations become closely associated with personal goal-setting behavior originating in our internalized system of values. Expectations are external in nature and are related to goals set collectively by society or by significant individuals in our lives. Assessment in this area is based on four components believed to be related to the development of positive self-esteem.

The first has to do with locus of control — whether one views personal success as dependent upon one's own efforts or external influences. Externally controlled individuals will tend to be more dependent on others and more willing to ride with the tide, accepting docilely things which happen to them. Internal individuals will more actively attempt to control self-destiny.

The second related concept is self-confidence — the feeling of self-worth and the belief that one is capable of handling things successfully. Those who lack self-confidence are often characterized as being timid, cautious, submissive individuals who feel inadequate, fearful, inferior and expect to be unsuccessful in dealing with new situations.

The third component is image in school settings. Those having favorable self-images are likely to experience subjective success with schoolwork, feel that they are favorably viewed and understood by teachers and enjoy class participation.

The final dimension considers how students feel about the quality of their relationships with others. Individuals who have difficulty in interpersonal relations will tend to believe that others have little confidence in or low regard for them.

GENERAL SCALE DESCRIPTION*

The self-esteem scale is comprised of 40 short, self-description statements. Twenty-one are positively worded—describing the student in a favorable light and 19 are negatively worded—characterizing the student in a negative vein.

Sample positively worded item: *I'm easy to get along with.*

Sample negatively worded item: *Things are all mixed up in my life.*
Response options available to the students are (1) very true of me, (2) mostly true of me, (3) mostly untrue of me and (4) very untrue of me.

The items within the scale are grouped to yield four subscale scores in addition to a total scale score.

Subscale 1: *Self-confidence* contains 10 items measuring feelings of success, self-determination, attractiveness and self-worth. Sample item: *I'm pretty sure of myself.*

Subscale 2: *Feelings of control over environment* contains 10 items tapping belief that success in school and work depend on effort, not luck. Sample item: *My getting good grades in school depends more on how the teacher feels about me than on how well I can do my work.*

Subscale 3: *Relationships with others* contains 10 items assessing the student's perceived ease in making and keeping friends and the student's feelings of acceptance by others. Sample item: *I often feel picked on by other kids.*

Subscale 4: *Self-image in school* comprises 10 items designed to measure feelings of success in school work, class recitation and teacher relationships. Sample item: *In class, I often feel 'put down' by teachers.*

NORM-REFERENCED SCORING

For norm-referenced scoring the item weighting scheme used is:

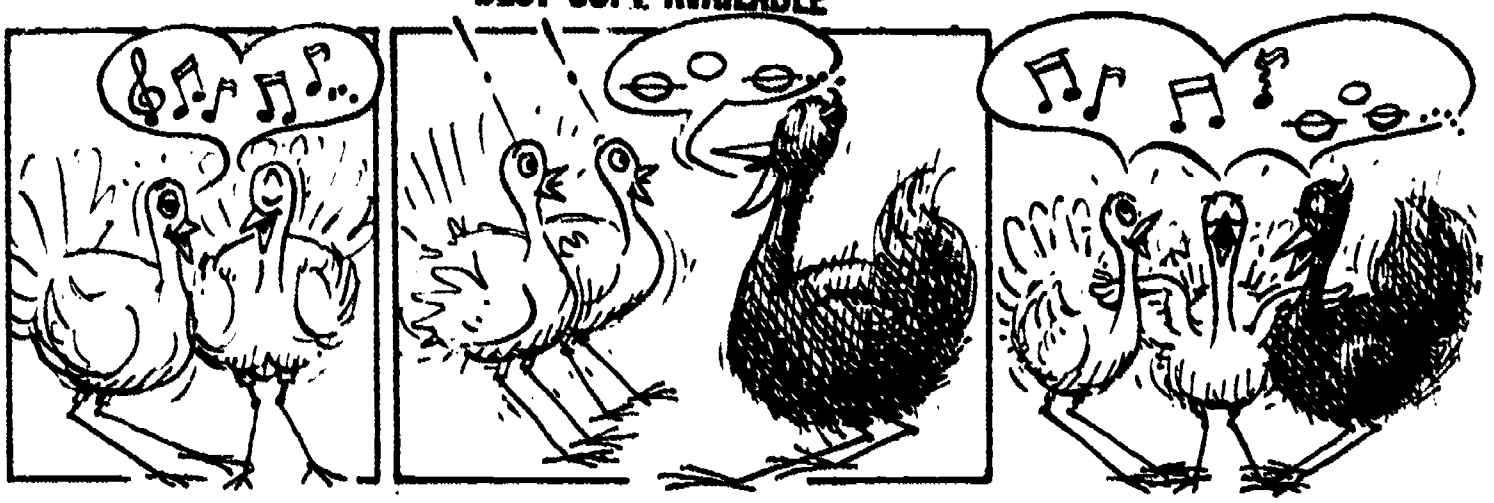
	Response Choices			
Item Direction	Very True of Me	Mostly True of Me	Mostly Untrue of Me	Very Untrue of Me
Positive	3	2	1	0
Negative	0	1	2	3

*The self-esteem scale is a result of extensive revision of the Goal I instrument which was used for grades 5 and 11. Richard L. Kohr and Nolan F. Russell from the Division of Educational Quality Assessment were responsible for the revisions.

CRITERION-REFERENCED SCORING

Responses are considered *favorable* if they reflect a positive self-image. An individual's score on a given scale (total or subscale) is the percentage of items to which a favorable response was given. For the self-esteem instrument the scoring scheme applied to the items is:

	Response Choices			
Item Direction	*Very True of Me*	*Mostly True of Me*	*Mostly Untrue of Me*	*Very Untrue of Me*
Positive	1	1	0	0
Negative	0	0	1	1



GOAL II Tolerance Toward Others

Quality education should help every child acquire understanding and appreciation of persons belonging to other social, cultural and ethnic groups.

GOAL RATIONALE

Students fulfilling the requirements of Goal II will more likely enjoy easy interaction with *all* people—speaking to and selecting as friends students of different origins and beliefs. They will be more willing to actively seek information or participation in activities which will increase their knowledge about different cultures and social settings.

The school experiences should help students learn to respect and interact easily with children who differ from them in various aspects (e.g., skin color, cultural traditions, economic status, religious beliefs, physical abilities, manner of speech and degree of intellectual competence).

MEASUREMENT RATIONALE

The processes and determinants of interpersonal interaction are complex, involving a myriad of perceptual, feeling and behavior responses.

The notion of tolerance toward others has meant different things to various theorists. Some define tolerance in terms of the *social distance* individuals keep between themselves and differing others. Others use tolerance to describe the tendency of individuals to prejudge or act toward differing others solely on the basis of the differing others' group membership.

The assessment of this goal area is based on still another component of tolerance. This component is the degree of comfort felt by individuals when in contact with differing others.

GENERAL SCALE DESCRIPTION*

Items describe situations where differing others interact with the individual. Differences are in terms of racial, religious and social background or physical and mental attributes. Twenty-nine items suggest an approach toward the student, e.g., *A cripple wants you to become a close friend.* Six items suggest an avoidance of the student, e.g., *A girl with a bad limp avoids you because she thinks you might make fun of her.* Response choices are I would feel (1) very comfortable, (2) comfortable, (3) slightly uncomfortable and (4) very uncomfortable.

*The tolerance toward others instrument was developed by Eugene W. Skiffington and Nolan F. Russell from the Division of Educational Quality Assessment and Peggy L. Stank and Tom McGinnis from the Division of Research.

The items within the scale are grouped to yield five subscale scores in addition to a total scale score. Assignment to subscales is based upon the characteristic of the hypothetical target person that makes that person different from the respondent. The five subscales are race, religion, socioeconomic status, intelligence and handicap. All subscales contain seven items.

NORM-REFERENCED SCORING

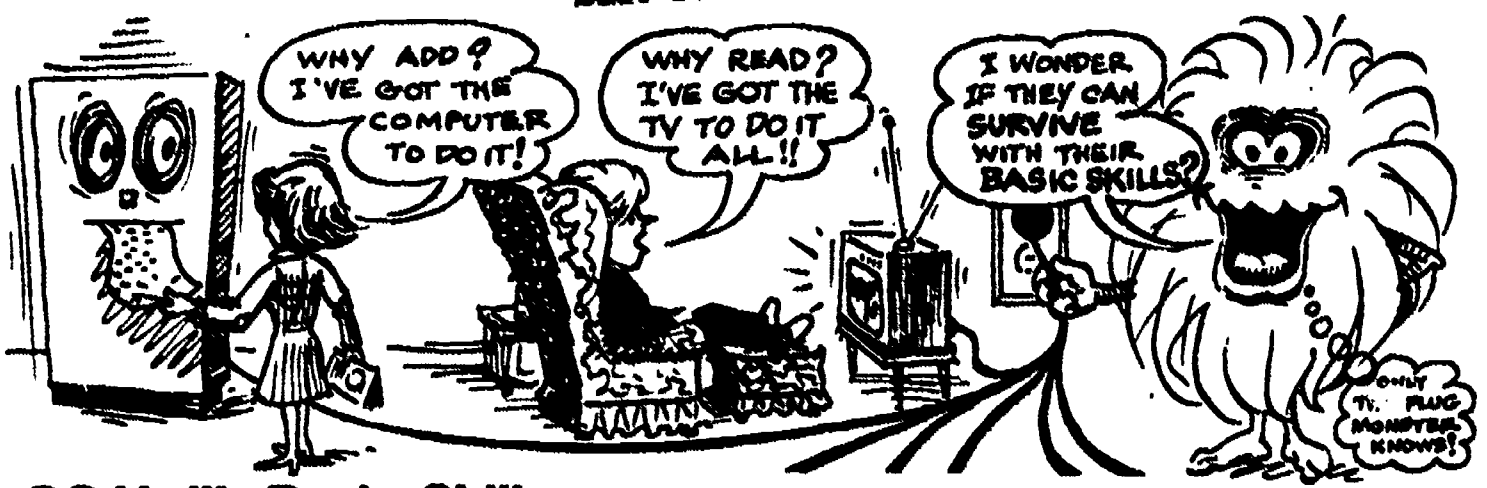
For norm-referenced scoring, the item weighting scheme is:

	Response Options			
Item Direction	Very Comfortable	Comfortable	Slightly Uncomfortable	Very Uncomfortable
Positive	3	2	1	0
Negative	0	1	2	3

CRITERION-REFERENCED SCORING

Responses are considered *favorable* if they reflect comfort when interacting with differing others or discomfort when being shunned by differing others. An individual's score on a given scale (total or subscale) is the percentage of items to which a favorable response was given. For the tolerance toward others instrument the scoring scheme for items is:

	Response Options			
Item Direction	Very Comfortable	Comfortable	Slightly Uncomfortable	Very Uncomfortable
Positive	1	1	0	0
Negative	0	0	1	1



GOAL III Basic Skills

Quality education should help every child acquire, to the fullest possible extent, mastery of the basic skills in the use of words and numbers.

GOAL RATIONALE

Mastery of the basic skills in the use of words and numbers is fundamental to achievement in all academic areas. Basic skills include the ability to get ideas through reading and listening, to handle mathematical operations, to reason logically and to respect evidence. The level of performance that can be reasonably expected in each of these areas will vary from school to school. However, it is of profound importance that the level of expectation in basic skills for any group of pupils shall not be underestimated or regarded as fixed.

MEASUREMENT RATIONALE

In 1969 when Pennsylvania's Educational Quality Assessment Program centered on 5th and 11th grade students, schools could select either of two standardized achievement batteries for measurement in this goal area. It quickly became apparent that the use of these tests increased the length of the testing time so as to cause great difficulty in scheduling and completing the entire questionnaire.

Therefore, the use of achievement batteries was discontinued and shorter verbal and math scales developed by Educational Testing Service which were group reliable were substituted.

In the verbal area assessment was directed at the ability to abstract or generalize and to think constructively, rather than at simple fluency or vocabulary recognition. The item type thought most appropriate was one using verbal analogies.

The test in the math area is directed at those mathematics skills and concepts all students should be familiar with and not skills and concepts attainable only by academically gifted persons.

GENERAL SCALE DESCRIPTION (VERBAL)*

The verbal scale contains 30 verbal analogies presented in a multiple-choice format. Each scale is timed (15 minutes). The scales are scored by giving one point for each correct answer. No adjustment is made for guessing.

Sample item: (Grade 8)

MEAL: BANQUET::
A diamond: jewel
*B car: limousine
C design: ornament
D silver: gold

GENERAL SCALE DESCRIPTION (MATH)**

The mathematics scale has 30 items and is timed (15 minutes). Its ability to identify specific strengths and weaknesses in math-related areas is limited. However, it is considered a good measure for the general level of math achievement on a group basis. Modern math concepts (set notion, modular arithmetic, etc.) and advanced concepts such as trigonometry, logic and geometric proofs are not included. Areas tapped are arithmetic computation, algebraic and geometric concepts and measurement. A multiple-choice format is used. Each item requires students to make a size comparison between two quantities. The scale is scored by giving one point for each correct answer. No adjustment is made for guessing.

Sample item: (Grade 8)

$$\frac{3}{9} = \frac{x}{27}$$

$$\frac{3}{7} = \frac{y}{21}$$

Column A	Column B
x	y

A. The part in Column A is greater.
B. The part in Column B is greater.
*C. The two parts are equal.
D. Not enough information is given to decide.

GOAL IV Interest in School and Learning

Quality education should help every child acquire a positive attitude toward the learning process.

GOAL RATIONALE

The school represents perhaps the most powerful single force in determining a person's overall attitude toward learning. The climate and learning atmosphere in the school, the educational experiences the school provides and the quality of the personal interactions it fosters between student and educator all shape the students' life-long attitudes toward learning.

The school experience should be such that students find the learning activities associated with it enjoyable and rewarding to the point that they are motivated to do well and to continue learning on their own initiative beyond the requirements of formal education. Everything possible should be done to ensure that the attitude of the teacher, the atmosphere of the school, and the school's physical condition contribute toward this end so that the individual--both as a child and later as an adult--will hold education high among his or her values.

MEASUREMENT RATIONALE

In assessing student feelings about education, it is necessary to examine more than just those feelings within the context of the students' present school experience. We must also determine how this experience is influencing the students' general future attitudes toward learning beyond the formal educational setting. The measurement device developed in support of this goal attempts to sample student attitudes in two areas: The first relates specifically to the present school experience while the second focuses on learning as a lifetime process.

GENERAL SCALE DESCRIPTION*

In this scale there are 30 statements about the school, teachers, course content and the learning experience. Fifteen items cast these areas in a favorable light, e.g., *Most of my subjects this year are worthwile.* The remaining items are negatively stated. e.g., *Teachers don't know what they are talking about.* Response options available to the student are (1) strongly agree, (2) agree, (3) uncertain, (4) disagree and (5) strongly disagree.

*The interest in school and learning scale is a result of extensive revisions of the Goal IV instrument which was used for grades 5 and 11. George E. Brehman from the Division of Research and Nolan F. Russell from the Division of Educational Quality Assessment were responsible for the revisions.

The items within the scale are grouped into two subscales each having 15 items.

Subscale 1:

Attitude toward learning measures the student's willingness to expend effort to learn and to appreciate the value of continued learning throughout life. Sample item: *It is very important to me to learn as much as I possibly can.*

Subscale 2:

Attitude toward school investigates the degree to which the student believes school attendance is important and the student's attitude toward the school setting, teachers and course work. Sample item: *Most of my classes this year are boring.*

NORM-REFERENCED SCORING

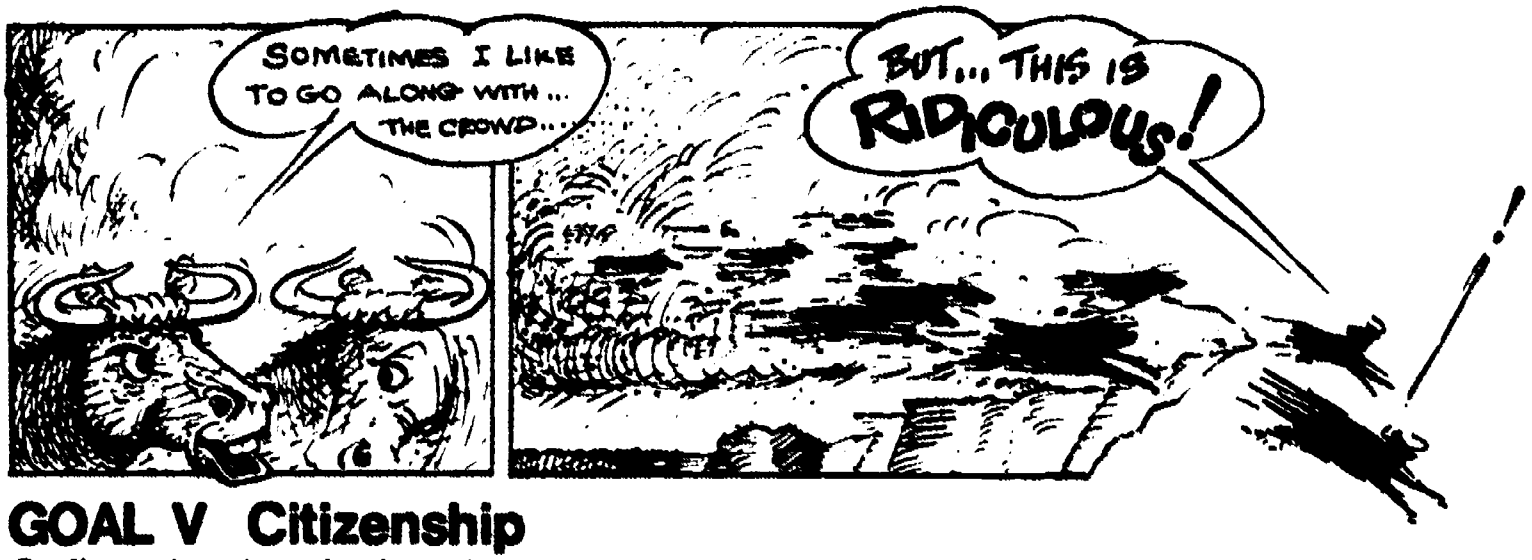
For norm-referenced scoring, the following weighting scheme is used:

	Response Options				
Item Direction	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
Positive	4	3	2	1	0
Negative	0	1	2	3	4

CRITERION-REFERENCED SCORING

Responses are considered favorable if they reflect student agreement with positive statements about school and learning or disagreement with negative statements concerning school and learning. A student's score on a given scale (total or subscale) is the percentage of items to which a favorable response was given. For this scale the scoring scheme applied to the items is:

	Response Options				
Item Direction	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
Positive	1	1	0	0	0
Negative	0	0	0	1	1



GOAL V Citizenship

Quality education should help every child acquire the habits and attitudes associated with responsible citizenship.

GOAL RATIONALE

Responsible citizenship embodies a much more complex concept than commonly expressed in love of country and participation in the democratic processes. Viewed in its broadest sense responsible citizenship implies a respect for law and proper authority, a willingness to assume responsibility for our own actions and for those of the groups to which we belong, respect for the rights of others and overall personal integrity.

Schools should encourage pupils to assume responsibility for their actions as well as the actions of the group. Opportunities should be provided for pupils to cooperate and work toward group goals and to demonstrate integrity in dealing with others. Pupils should be given the chance to take the initiative and assume leadership for group action as well as lend support to group efforts as followers.

MEASUREMENT RATIONALE

The mores, codes, laws and social expectations of society provide the reference points for judging which behaviors reflect *responsible* citizenship and which indicate *poor* citizenship. A review of literature revealed that the National Assessment of Educational Progress developed nine general citizenship objectives. The criterion for inclusion of any one objective was its relative importance to society as agreed upon by a committee of scholars and lay people.

These national objectives were used to provide the frame of reference for what was to be measured. Objectives in the factual domain such as (a) knowing structure of government and (b) understanding problems of international relations were not considered in developing the scale.

Arriving at a satisfactory definition of citizenship was much less complicated than applying the definition to the assessment of students' attitudes and behaviors. The display of responsible citizenship behaviors like *honesty* or *integrity* are most often situational.

A student's display of good citizenship behavior under one set of motivating conditions tells us little about the way he or she can be expected to behave if those conditions are altered. The context in which the behavior is elicited therefore becomes at least as important in determining the outcome as the predisposition of the individual involved.

To assess citizenship, a behavior-referenced model incorporating elements related to the psychological notion of *threshold* is used. In reference to citizenship, threshold refers to that set of conditions necessary to bring about the desirable responses. Thus by varying the situation and introducing conditions of reward and punishment we are able to determine the cutoff levels at which the student will display positive behavior. In this way it is possible to assess not only the students' predisposition to behave in a manner consistent with responsible citizenship but also to provide some measure of the intensity of that predisposition across a wide spectrum of situations.

GENERAL SCALE DESCRIPTION*

Fifty-four items measure willingness to exhibit good citizenship in many social situations under a variety of motivating conditions. Social contexts are given by 18 stories, each posing a problem and suggesting an action predefined as good or poor citizenship. Each story has three items which list positive or negative consequences resulting from the action. Students are asked to decide whether to take the action for each consequence.

Sample Story (grade 8): Morton has broken a school window but did not mean to. If I were Morton, I would TELL THE PRINCIPAL OR TEACHER about my breaking the window when I knew. . .

Sample item set:

		Yes	Maybe	No
1.	The principal would make me stay after school.	Y	M	N
2.	My parents would have to pay for the window.	Y	M	N
3.	I would have to pay for the window.	Y	M	N

The items within the scale are grouped in such a way as to yield three subscale scores in addition to a total score.

Subscale 1: *Concern for the welfare and dignity of others* contains 15 items (item sets from five stories) designed to measure concern for the feelings of others, willingness to protest unjust treatment of others, and the tendency to accept new people into a group. Also measured is the degree of restraint from teasing or degrading others.

Subscale 2: *Respect for law and authority* has 21 items measuring the willingness to report law-breaking of others, obey authorities during emergencies and prevent classroom disruptions. Also assessed is the degree of restraint from violent actions that could harm others or damage property.

Subscale 3: *Personal responsibility and integrity* has 18 items which tap the willingness to honor

*Nolan F. Russell from the Division of Educational Quality Assessment is the author of the citizenship scale.

self-made commitments to individuals or groups and the willingness to take responsibility for one's own mistakes and to report mistakes made in one's favor.

NORM-REFERENCED SCORING

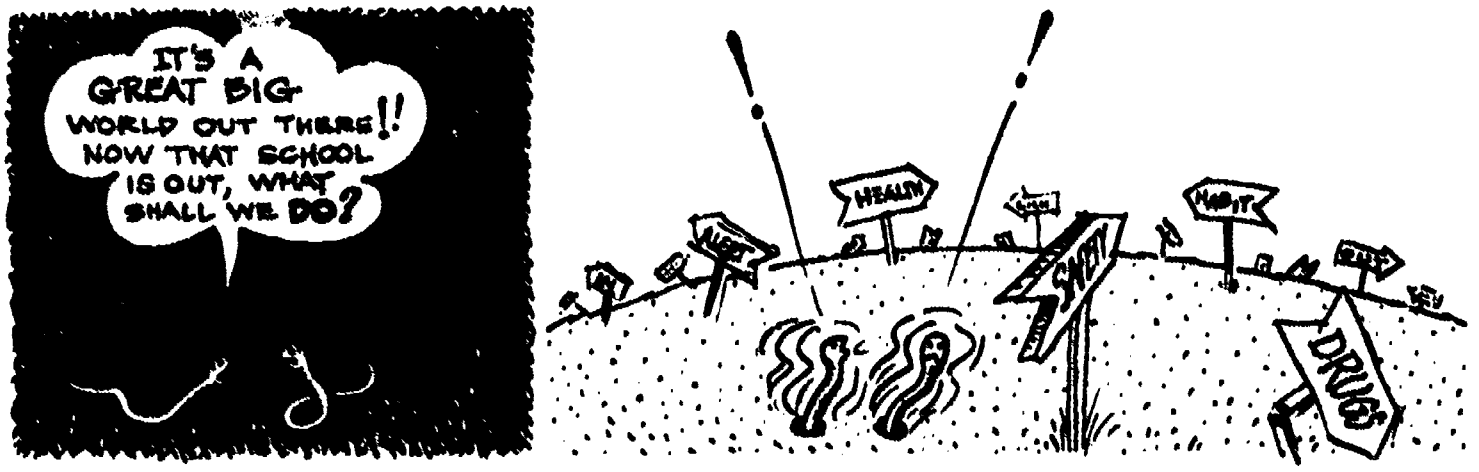
The following item weighting scheme is used for norm-referenced scoring:

	Response Options		
Behavior Direction	*Yes*	*Maybe*	*No*
Positive Citizenship	2	1	0
Negative Citizenship	0	1	2

CRITERION-REFERENCED SCORING

Responses are considered *favorable* when they reflect a willingness to display proper citizenship behaviors or an unwillingness to use poor citizenship behaviors. A student's score on a given scale (total or subscale) is the percentage of items to which a favorable response was given. For the citizenship scale the scoring scheme applied to the items is:

	Response Options		
Behavior Direction	*Yes*	*Maybe*	*No*
Positive Citizenship	1	0	0
Negative Citizenship	0	0	1



GOAL VI Health Habits

Quality education should help every child acquire good health habits and an understanding of the conditions necessary for maintaining of physical and emotional well-being.

GOAL RATIONALE

In their own interest, as well as in the interest of society at large, children should know how to take care of themselves and how to keep physically fit. They should know what the requirements are for physical and mental health and what practices, harmful to health, should be avoided. After gaining this knowledge they should acquire habits which increase the probability of remaining healthy and fit throughout life.

In cases where the home has been deficient in encouraging the child to practice sound health habits, the school has an obligation to be aware of the situation and to see that opportunities to remedy the deficiency are provided.

MEASUREMENT RATIONALE

Understanding how diseases and their prevention, dental care, nutrition, personal hygiene, safety and drug use relate to the structure and function of the human body is an important first step in each individual's willingness to consistently exhibit habits which are conducive to the maintenance of personal health and well-being. One does not need to be a doctor to display good health practices or a lawyer to display good citizen behaviors.

Therefore, assessment in this goal area attempts to get at students' willingness to display proper health behaviors in a variety of situations.

GENERAL SCALE DESCRIPTION*

The scaling technique in this inventory is similar to the psychophysical method of limits. This method holds the behavior constant while systematically allowing the stimuli to vary. The strength of the stimulus (in physical units) which is required to cause a change in the behavior is used to define the threshold of that behavior.

In the case of this health behavior inventory, the student is asked to decide whether he or she would take a given health-related action. Each action is predefined as indicating either

*The health habits scale was developed jointly by Nolan F. Russell and Eugene W. Skiffington from the Division of Educational Quality Assessment and Mary L. Lydon from the Division of Research.

good or poor health practice. Stimulus contexts surrounding the choices are systematically varied. The health-behavior threshold is defined in terms of the severity of the stimulus contexts tolerated before changing from good to poor health behavior. More specifically performance is used to infer health-behavior threshold by identifying the supportive contingencies in the environment necessary to maintain good health practices.

The format of each question is a situational story about a make-believe junior high student. The respondent is first asked to consider taking a specific action. In each question three motivation-including conditions, i.e., rewards and punishments, are made contingent upon the taking of the action.

Sample story: When Norma had the flu the doctor gave her some medicine. The medicine also took away the stomach ache Norma had. After she got over the flu, Norma had another stomach ache. If I were Norma, I would TAKE THE MEDICINE AGAIN when I thought. . .

Sample item set:	Yes	Maybe	No
1. The medicine tasted good.	Y	M	N
2. It might cure my stomach ache quickly.	Y	M	N
3. My parents might *not* want me to take it.	Y	M	N

The items within the scale were grouped to yield three subscale scores in addition to a total scale score.

Subscale 1: *Personal and community health* contains 21 items (seven health situations). Content includes willingness to follow proper diet, to take proper medical precautions, to use good personal hygiene practices and to refrain from interpersonal contacts when ill.

Subscale 2: *Personal and community safety* contains 18 items from six health situations. Measured is the degree of restraint from unnecessary risk-taking at home, at school and at play and restraint from submitting others to undue risks.

Subscale 3: *Drugs* contain five situations with 15 questions to measure restraint from (1) improper use of prescription drugs, (2) experimentation with drugs and (3) maintaining close contact with others who are using drugs. Improper use of prescription drugs includes restraint from using old medicine, medication prescribed for others, or more medicine than has been prescribed by the doctor.

NORM-REFERENCED SCORING

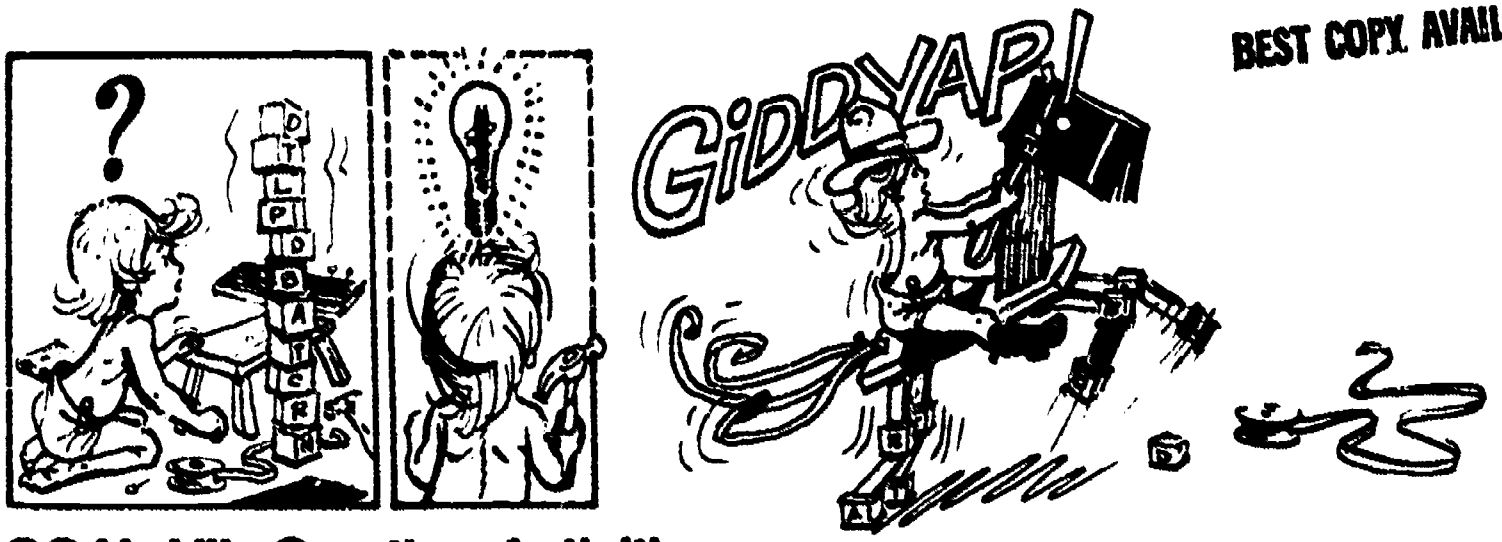
For norm-reference scores, the following item weighting scheme is used:

	Response Options		
Behavior Direction	*Yes*	*Maybe*	*No*
Positive Health Behavior	2	1	0
Negative Health Behavior	0	1	2

CRITERION-REFERENCED SCORING

Responses are considered *favorable* when they reflect a willingness to take good health-related actions or an unwillingness to display behaviors that might be harmful to health. A student's score on a given scale (total or subscale) is the percentage of items to which a favorable response was given. For the health scale the scoring scheme applied to the items is:

	Response Options		
Behavior Direction	*Yes*	*Maybe*	*No*
Good Health Practice	1	0	0
Poor Health Practice	0	0	1



GOAL VII Creative Activities

Quality education should give every child opportunity and encouragement to be creative in one or more fields of endeavor.

GOAL RATIONALE

The notion of creativity has been variously defined. It is used here to encompass worthwhile activities that children initiate and pursue on their own-- activities having an outcome that is perceived by the children themselves or by others as a contribution to their world. Such activities can be found in a wide variety of fields, not only the sciences and the arts, but also the organization of human affairs and the development and exercise of salable skills in the production of practical things that enrich our way of living.

The school environment should encourage and reinforce activities that can enable children to express themselves creatively and productively.

MEASUREMENT RATIONALE

Attempts to assess creativity have traditionally used methods which analyze the components of the creative process or subjective judgments about the quality of the product of the creative act. Neither of these procedures is particularly well adapted to a large scale assessment effort which covers the broad spectrum of creative talent represented in the school. In order to overcome this problem a two-dimensional model of creativity was proposed which provided a theoretical basis for the assessment of Goal VII. The first dimension is based on the student-expressed interest in participating in creative activities, while the second attempts to determine the extent of recognition gained through active involvement. This approach seems sound since the Goal VII statement stresses opportunities and encouragement for all students relative to creativity rather than emphasizing individual talent and production in any one area.

GENERAL SCALE DESCRIPTION*

The creative activities checklist contains 36 activities which require originality in visual arts, performing arts, science and writing. Sample activities include performing an original scientific experiment with living things, writing an original poem, modeling an outift in an original way, performing an original magic or novelty act.

Response options give six ways to show degree of involvement in each activity. Options are (1) No, and have not wanted to; (2) No, but have wanted to; (3) Yes, but with no recognition;

*The creativity scale was developed by James F. Hertzog and Nolan F. Russell, both from the Division of Educational Quality Assessment.

(4) Yes, with teacher or adult leader recognition; (5) Yes, with school-wide recognition; and (6) Yes, with area-wide recognition. The scale contains four subscales each having nine items.

Subscale 1:

Visual arts contains nine items, some dealing with more than one activity. Activities include sculpturing; cartooning; printmaking; graphic design; painting; photography; flower arrangement; design of window displays, stage sets, decorative items and clothing.

Subscale 2:

Performing arts contains nine items which include activities dealing with singing, speech, music, magic, modeling, directing, acting and sports.

Subscale 3:

Writing arts has nine items related to writing such as poetry, news, essays, stories, scripts, letters, jokes and recipes.

Subscale 4:

Science activities contains nine items such as performing experiments using physical objects or living things, constructing models to show a scientific principle, exploring, training animals, directing recreational activities, developing campaign strategies for (school) elections, working with radios or other electronic equipment and designing gadgets.

NORM-REFERENCED SCORING

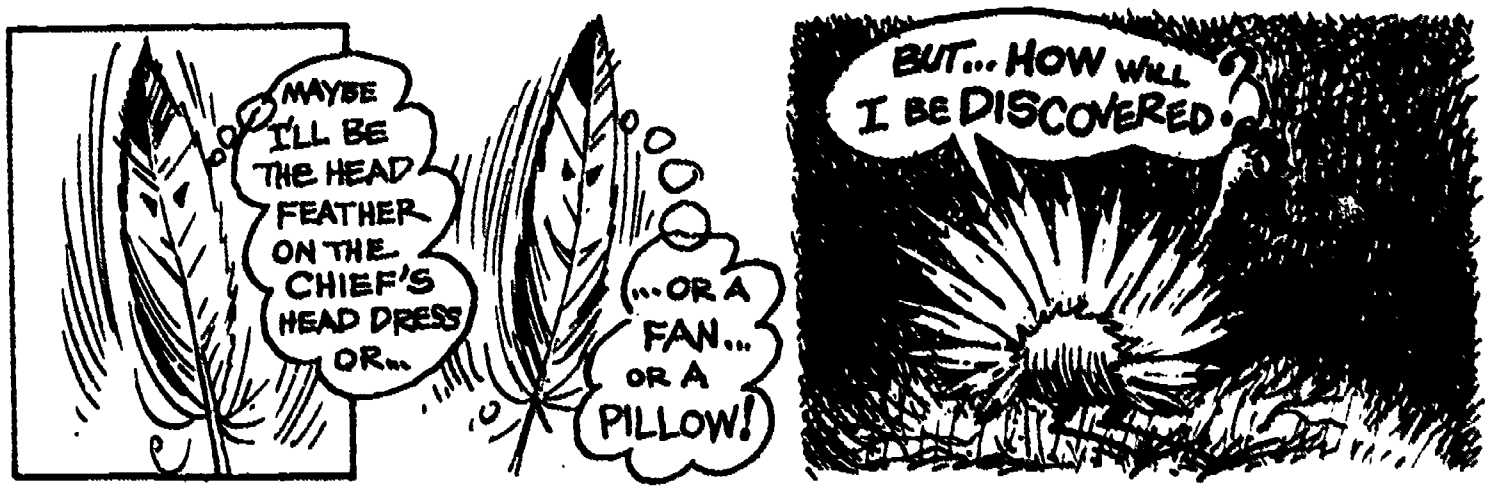
All items in this scale are positively worded. Each item describes an activity and asks the students to describe the level of their involvement in that activity.

Response Options	*Score Obtained*
(1) No, and have not wanted to	0
(2) No, but have wanted to	1
(3) Yes, but no recognition	2
(4) Yes, with teacher or adult leader recognition	3
(5) Yes, with school-wide recognition	4
(6) Yes, with area-wide recognition	5

CRITERION-REFERENCED SCORING

Two criterion-referenced scoring methods are used for this scale. The first defines as *favorable* those choices which reflect a willingness to try the activities presented in the scale. Thus only the option *No, and have not wanted to* is considered unfavorable. Scores generated from this method are called Attitude Toward Creative Activities.

The second scheme defines as *favorable* those choices indicating that the student has actually participated in the activity. Thus two choices are considered unfavorable: *No, and have not wanted to* and *No, but have wanted to.*



GOAL VIII Vocational Knowledge

Quality education should help every child understand the opportunities open to him or her to prepare for a productive life and help each child to take full advantage of these opportunities.

GOAL RATIONALE

Students should be aware of the vast array of possibilities for continuing self-development in the world of work so that they will be motivated to pursue excellence in all forms of human endeavor that are appropriate for them individually.

Most children can profit from some form of education beyond high school, whether it be a four-year college, a school of nursing, a community college, a technical institute or the like. Each student should be aware of these opportunities and seek out the kind of education best suited to his or her talents and interests. This goal also implies that the school will provide students with guidance that will enable them to do so.

MEASUREMENT RATIONALE

Vocational development, for purposes of assessment, is a series of processes involving both the acquisition of knowledge about different kinds of work and the forming of attitudes which will enhance one's chances of succeeding in the work-a-day world.

In the initial stages of vocational maturity students become aware of different kinds of work and workers. This is followed by a growing understanding of the relatedness of educational and occupational opportunities. The more vocationally mature students will actively seek information, accept personal responsibility for career decisions and finally base their career choices upon a realistic appraisal of their interests, achievements and aptitudes.

The first stage, measured at this grade level, assumes that the awareness of jobs and job opportunities will lead to development of vocational maturity.

GENERAL SCALE DESCRIPTION*

This scale contains 30 items which focus on the student's knowledge concerning the duties, conditions, training, salary and educational requirements of various occupations. The multiple-choice format requires the student to select the best answer from alternatives. One point is given for each correct response. No adjustment is made for guessing.

*The authors of this scale are Francis J. Reardon and James P. Lewis from the Division of Research.

Sample items:

Which of these jobs requires a period of apprenticeship?

*A. electrician
B. stockbroker
C. porter
D. truck driver

A claims adjuster would most likely be working for:

*A. an insurance company
B. a hospital
C. a welfare office
D. a department store

GOAL IX Appreciation of Human Accomplishments

Quality education should help every child to understand and appreciate as much as possible of human achievement in the natural sciences, the social sciences and the humanities and the arts.

GOAL RATIONALE

Students should be encouraged and helped to gain knowledge about human accomplishments. Possessing knowledge they will then be ready to receive and not to avoid the stimuli that the sciences and arts provide. At the next level, they will be ready to more clearly and consciously perceive these stimuli and will begin to discriminate among art forms. When they reach the next stage of development, they will be ready to respond rather than merely attend to phenomena–they will choose to see a play, to read of a famous scientist or to contemplate the design of a building.

Insofar as possible the school experience should provide an increasing openness to the life of the mind and an increasing ability to find meaning for one's own life in the heritage of the past and in the intellectual thrusts of the present age.

MEASUREMENT RATIONALE

Attitudes associated with the understanding and appreciation of human accomplishments may be inferred from samplings of behavior taken at several points along a response hierarchy. The lowest point in the hierarchy is represented by behaviors indicating a state of passive receptivity reflecting little more than an awareness that certain human endeavors exist. At the highest point of this hierarchy are overt behaviors resulting in direct involvement in the activities and inferring high motivation. Between these two extremes are several intermediate steps based on the *value* placed on the activities and willingness to *receive* stimuli that these activities provide.

In developing the assessment model to be used in this goal area, it was determined that the instrument would not attempt to sample behaviors at either extreme. Instead items were designed to concentrate on attitudes concerned with the degree of value placed by students on various areas of human accomplishment and the willingness of students to seek out environments where firsthand experience in these endeavors would be possible.

GENERAL SCALE DESCRIPTION*

This scale contains 48 items measuring how much value the students place on human

*The appreciating human accomplishments scale was developed by Joan S. Beers and Nolan F. Russell for the Department of Education.

achievements in the arts and sciences and the degree to which they are willing to vicariously receive stimuli from these endeavors. Areas included are literature, art, athletics, ecology, government, science, music and drama. The scale is organized into two subscales each having 24 items.

Subscale 1:

Valuing measures the amount of importance the student attaches to achievements in the arts and sciences and how much the student values the role played by people in these areas. Sample item: *Most scientists don't care how their work affects people.*

Subscale 2:

Receiving measures willingness to learn more about achievements in the arts and sciences and to seek out firsthand information on what people in these areas are doing. Sample item: *It would be fun to watch people paint at an art studio.*

NORM-REFERENCED SCORING

The item weighting scheme used for norm-referenced scoring is:

	Response Options		
Item Direction	*Agree*	*Uncertain*	*Disagree*
Positive	2	1	0
Negative	0	1	2

CRITERION REFERENCED SCORING

Responses to this scale are considered favorable when they reflect agreement with statements which (1) stress the value of human endeavors in the arts, sciences, politics, etc., or (2) suggest that it is personally rewarding to approach the people and places associated with these endeavors. For the appreciation of human accomplishments scale the following scoring scheme is used.

Item Direction	*Agree*	*Uncertain*	*Disagree*
Positive Statements	1	0	0
Negative Statements	0	0	1

GOAL X Coping with Change

Quality education should help every child to prepare for a world of rapid change and unforeseeable demands in which continuing education throughout adult life should be a normal expectation.

GOAL RATIONALE

Ability to cope with a rapidly changing world is important for today's youth. The development of the abilities and their associated attitudes which allow the individual to view change as an opportunity rather than a threat poses a new challenge for education.

Schools should help students develop attitudes of openness to the possibilities of change – change in their personal world as well as external change. Students should be encouraged to show tolerance for uncertainty and to welcome new experiences.

MEASUREMENT RATIONALE

Ability to cope with change and deal effectively with frustration is essential to personal adjustment. These adaptive behaviors are seldom learned in response to external changes of great magnitude and import but are acquired as part of a gradual process requiring daily changes in the life of the student.

Assessment in this goal area attempts to draw upon several elements believed to be associated with a student's ability to accommodate change and to adapt emotionally and behaviorally to unexpected or sudden alterations in the environment. Primary among these are measures of the student's ability to tolerate frustration and uncertainty and to apply past learnings and coping behaviors in new and different situations.

The situations presented as a means of measuring these attitudes and behavioral dimensions were gleaned from student responses to open-ended questions asking for descriptions of events they had experienced which necessitated some form of adjustment and which were remembered as being difficult to cope with.

GENERAL SCALE DESCRIPTION*

Thirty-five items measure emotional and behavioral reactions to change. The scale's format contains seven stories describing unpleasant change situations in which student's expectations or needs are not met. These situations were obtained from previous student statements describing events that were difficult to adjust to. Five reactions predefined as indicating positive

* The preparing for a changing world scale was authored by Nolan F. Russell, Division of Educational Quality Assessment.

or negative adaptation to change are given following each story. The purpose of the scale is to get at student reactions in response to a variety of events, not to predict what students will do in the particular situations presented.

Sample situation: I was elected class president. I came home to tell my parents the good news. They told me that my dad had taken a job out of state and we were going to move in two weeks. So I had to withdraw from school and move.

Sample items: If this happened to you, how much *time* would you spend on each thing listed below:

	A Great Deal of Time	*Some Time*	*Very Little Time*	*No Time*
1. Being upset.				
2. Trying to find someone to stay with so I could remain in my school.				
3. Planning a going-away party.				
4. Fighting with my parents.				
5. Reading about the place we are going to move to.				

Subscale 1: *Effective solutions* contains 13 items to measure the tendency to try solutions reflecting positive adjustment to change. In the above, sample items three and five are assigned to this subscale.

Subscale 2: *Ineffective solutions* contains 13 items to measure tendency to avoid use of aggressive or withdrawing reactions in face of change. In the above, sample items two and four are assigned to this subscale.

Subscale 3: *Emotional adjustment* contains nine items to measure the perception of the length of time needed for the student to adjust emotionally to change. Item one above is assigned to this subscale.

NORM-REFERENCED SCORING

The item weighting scheme for norm-referenced scoring is:

	Response Options			
Type of Items	*A Great Deal of Time*	*Some Time*	*Little Time*	*No Time*
Effective Solutions	3	2	1	0
Ineffective Solutions	0	1	2	3
Emotional Adjustment	0	1	2	3

CRITERION-REFERENCED SCORING

Responses are considered favorable when they reflect (1) a willingness to adjust positively, (2) an unwillingness to withdraw or become aggressive and (3) a rapid emotional adjustment to change. An individual's score on a given scale (total or subscale) is the percentage of items to which a favorable response was given. The item weighting scheme for the *preparing for a changing world* scale is:

	Response Choices			
Item Type	*A Great Deal of Time*	*Some Time*	*Very Little Time*	*No Time*
Effective Solution	1	1	0	0
Ineffective Solution	0	0	1	1
Emotional Adjustment	0	0	1	1

35

SECTION TWO
Measurement Safeguards

THE OVERALL PICTURE

During the first two weeks of March 1974, 53,226 students attending 240 Commonwealth intermediate schools completed the Grade 8 EQA Inventory. This represents nearly a quarter of a million hours of student time. Do the outcomes which are summarized by the *Manual for Interpreting Intermediate School Reports* accurately reflect student progress on the 10 state education goals? To answer this question one must know where the tests came from, what they really measure, how accurately and reliably they measure it, and how much influence faking and response bias have on the final results. The following two sections highlight the safeguards used to produce high quality instruments and show how the instruments stood up to these checks.

A WORD ABOUT ATTITUDES

Attitudes, beliefs, values, etc., are abstractions. Nevertheless they are real enough to each individual holding them. They are typically thought of as a state of readiness--a predisposition to act or react in a certain way when faced with certain situations. A person's attitudes are always present but remain dormant most of the time. They are expressed in speech or other behavior only when the object of the attitude is perceived. A person may have strong attitudes for or against astrology but actively express them only when some issue connected with astrology arises--or when confronted by an attitude scale! Attitudes are often reinforced by beliefs (the cognitive component) and attract strong feelings (the emotional component) that will lead to particular behaviors (the action tendency component).

The measurement of attitudes always involves making inferences. Since the attitudes cannot be seen or measured directly, we must infer their presence from consistencies that appear in the individual's behavior. Observing individuals across time in everyday situations is probably the best way to learn how the individual thinks, feels and acts.

Clearly, this method is much too cumbersome and costly when we want to investigate the intensity and direction of attitudes for a large number of people, forcing us to rely instead on verbal reports of the individuals concerned.

WHAT ABOUT PAPER AND PENCIL TESTS OF ATTITUDES?

The use of paper and pencil techniques for measuring attitudes is often questioned. These questions are directed at both the test and the test taker. Test critics are concerned with the possibilities that:

1. People misunderstand what the items are asking.
2. People don't always tell the truth on this type of test.
3. Scores on tests of this type are seldom presented in any meaningful way.
4. People might not respond consistently to similar questions.
5. People might respond to items differently at various times.

In the development of the EQAI scales, these and other concerns were taken into consideration.

From the outset, all pilot instruments were put through an obstacle course of checks and balances designed to determine their susceptibility to various errors of measurement. The tests were then revised and submitted to additional field trials. This philosophy of *testing the tests before using them to test people* resulted in a five-year developmental period requiring strong cooperation between the Department of Education and over 175 Commonwealth school districts.

YES, BUT CAN THEY READ IT?

If students can't read the tests, the tests can't *read* the students. Although it is impossible to control the range of verbal comprehension in a program assessing over 50,000 students, it is feasible to develop scales that can be read and understood by the vast majority of respondents.

Toward this end, monitors conducting field-trials were required to submit detailed reports on the understandability of directions, item sentence structure and vocabulary for each instrument. Also, students were asked to comment on the scales and identify words and/or items that didn't make sense to them. Changes resulting from these data increased the readability of all EQA instruments.

After each scale was finalized, its readability level index was estimated by the Gunning-Fog formula. This formula takes into consideration both the average number of sentences and the percentage of three-or-more syllable words contained in 100 words. The index is expressed in *grade level* terms. An instrument with a readability index of 5.0 should be understood by the average student just entering 5th grade. Nine scales within the inventory have Gunning-Fog readability indices ranging from 4.0 through 5.0. The creative activities and appreciating human accomplishment scales have readabilities of 6.0 and 6.7 respectively.

Results from a study involving a limited number of special education students in a large suburban school district indicate that the majority of the tests comprising the battery need additional revision before becoming appropriate for our special education population.

FROM ITEM CONTENT TO ANSWER SHEET

Handing someone a list of attitude questions, waiting 10 minutes, then collecting it does not insure that the check marks you find in the answer column were made solely in response to the content of statements on the list. Those *other factors* influencing answers are called response sets. The two most common types of response set contamination are the tendency of respondents to tell you what they think you want to hear (make a good impression) or to randomly check answers without regard to item content.

In their developmental stages, all scales and items were checked against a special 36-item instrument designed to pick up the tendency to make socially desirable answers. This scale* contained such questions as:

(1) *I never forget to say please and thank you.*

(2) *Sometimes I don't like to obey my parents.*

(3) *I always finish my homework on time.*

(4) *Sometimes I do things I've been told not to do.*

Those items and scales whose scores were found to be associated with socially desirable responding were deleted or revised to minimize the relationships found. The correlations between the final total instrument scores and a 30-item version of the social desirability scale range from .04 through .08 (N=400).

We have tried to minimize the effects of position bias by including both positively and negatively worded items in the goal scales. The creative activities check-list is the only EQAI

*Item sources for this instrument included the Crown-Marlow Scale (1960) and the Childrens' Social Desirability Scale by Crandall (1965).

instrument which does not contain reflected items. An estimate of response position bias can be determined by counting the number of *perfect scores*. To get a perfect score the student must answer *Yes, I have done this activity and have received area-wide recognition for its quality* to each of 36 separate activities--a highly questionable feat! Fewer than one-half of one per cent of the 50,000 respondents obtained a perfect score on the scale.

FROM ITEM ANSWERS TO TEST SCORES

To give school personnel a clearer picture about the performance of their students on the EQAI, two scoring methods are used for each attitude scale. The first scoring method organizes the response options with each scale into a hierarchy. Different scoring weights are then applied to each level of the hierarchy. Consequently, for an item such as *I like school* with response options (a) strongly disagree, (b) disagree, (c) agree and (d) strongly agree, weights of zero through four are applied respectively to each answer choice. This method is based on the assumption that *strongly agreeing* with a statement is more positive than merely *agreeing* with the statement.

Item scores obtained by this method are summed and used to give norm-referenced information about student performance. How well a group of students perform on the scales is determined by the relationship of their scores to other student groups. Thus *norm-performance* tells very little about favorableness of student responses.

To obtain information about favorable and unfavorable responding, a criterion-referenced scheme is used. This scoring method is based on the notion that each item within the scale offers the respondent the chance to show a positive or negative attitude toward the specific content presented by the item. Hence, the response choices to the above item are scored by assigning a one to both the *strongly agree* and *agree* choices and a zero to the remaining choices.

The number of positive responses given by each student is compared to an independently determined *standard* or *criterion*. If the number of favorable responses meets or exceeds the standard, the student is said to have *achieved* the standard. In the case of scales used in EQAI, three criteria were set: Level One requires students to respond favorably to more than 35 per cent of the items; Level Two requires favorable responses to more than 50 per cent of the items. Level Three requires favorable responses to more than 70 per cent of the items.

RESPONSE CONSISTENCY WITHIN THE TESTS

Reliability is that characteristic of a measuring instrument which deals with consistency of results--either within the scale itself (internal consistency) or over time (stability). Reliability coefficients are reported as two-place decimal figures ranging from .00 to 1.00. As the instrument increases in reliability the coefficient increases in value.

Reliability coefficients are interpreted as the proportion of the variance in a set of scores which is caused by variation in the examinees true scores, rather than by errors of measurement.

The coefficients are derived by taking into account the length of the test and the extent to which test items contribute mutually confirming or consistent information.

The KR-20 reliability formula is used for the knowledge scales scored on a *right vs. wrong* basis. For the attitude scales, coefficient alphas give us estimates of scale and subscale internal consistency. Internal-consistency reliabilities based on criterion-referenced scoring of the scales are obtained using Livingston's formula.* As the magnitudes of these coefficients increase,

*Livingston, Samuel A., Criterion-Referenced Applications of Classical Test Theory *Journal of Educational Measurement*, Vol. 9, No. 1, Spring 1972, pp 13-25.

we can be more confident that errors of measurement are unlikely to make a difference between meeting or not meeting the criterion for many of the examinees.

Table 1 presents seven separate internal consistency estimates for each sub and total EQA scale. These are based upon a sample of 3,500 student records randomly drawn from all 240 schools administering the inventory in March 1974. Therefore, these reliability findings can be generalized across various schools, communities and test settings within the Commonwealth. Sub and total scale names are presented in acronym form. For complete names refer to Section One.

To clarify test appropriateness for students of differing reading and achievement levels, coefficient alphas are given for low achievement (N=1160), average achievement (N=1180) and high achievement (N=1080) students groups. These groups are defined by scores on the composite math-verbal scale.

As a rule of thumb, Kelley* has proposed that tests designed to discriminate between groups should display reliabilities greater than .50. Note that only three subscales (CONTENV, SES, and INTELL) fail to meet this minimum criterion and then only for the low ability student grouping. Also evident from Table 1 is the increase in reliability for most scales as ability level increases and the high reliability obtained on all total scale scores.

The three extreme right columns of Table 1 show criterion score reliabilities of 3,500 records across the three criteria levels. Here again the reliabilities are very high. This indicates that the tests are capable of eliciting consistent responses from students.

Other indicators of internal consistency available in printout form at the Division of Educational Quality Assessment in Harrisburg are ratios between standard-error of measurement and standard deviations, average inter-item correlations and items-to-total correlations for all sub and total scales. These indicators confirm conclusions obtainable from Table 1.

---

* T. L. Kelley, *Interpretation of Educational Measurements*, New York: Harcourt, Brace, and World, Inc. 1972.

TABLE 1

Selected Internal Consistency Reliability Estimates on Sub and Total EQA Instruments

Scale Names	Number Of Items	Norm Scoring Ability Group				Criterion Scoring (Total Group)		
		Low	Middle	High	Total	Level One	Level Two	Level Three
CSELF	10	.68	.77	.81	.76	.94	.75	.72
CONTENV	10	.49	.60	.67	.61	.96	.72	.58
RELATE	10	.73	.81	.85	.80	.98	.82	.75
SCHLIMAG	10	.65	.75	.80	.75	.85	.72	.79
TOTAL SCALE	40	.83	.89	.91	.88	.99	.93	.87
RACE	7	.63	.73	.79	.72	.79	.64	.72
RELIG	7	.66	.76	.81	.75	.80	.70	.78
SES	7	.42	.57	.65	.55	.85	.60	.54
INTELL	7	.42	.55	.61	.52	.61	.45	.72
HANDCP	7	.58	.69	.73	.66	.87	.66	.58
TOTAL SCALE	35	.81	.88	.90	.87	.96	.87	.88
GOAL - 3V	30	N.A.	N.A.	N.A.	.80	N.A.	N.A.	N.A.
GOAL - 3M	30	N.A.	N.A.	N.A.	.82	N.A.	N.A.	N.A.
LEARN	15	.81	.81	.84	.82	.94	.84	.80
SCHOOL	15	.87	.88	.89	.89	.90	.84	.92
TOTAL SCALE	30	.90	.91	.92	.91	.96	.90	.91
WELFDIG	15	.68	.83	.85	.81	.79	.84	.97
LAWAUTH	21	.75	.88	.90	.86	.84	.90	.98
RESPINTG	18	.75	.86	.88	.84	.81	.90	.98
TOTAL SCALE	54	.87	.94	.94	.93	.92	.94	.99

PHEALTH	24	.64	.78	.93	.78	.80	.95	.84
SAFETY	15	.76	.85	.85	.85	.84	.98	.88
DRUGS	15	.75	.82	.89	.83	.93	.89	.86
TOTAL SCALE	54	.83	.90	.87	.90	.94	.97	.91
VISLART	9	.75	.72	.76	.75	.79	.89	.74
PERFMART	9	.79	.77	.73	.77	.77	.95	.78
SCIENCE	9	.75	.76	.74	.75	.88	.80	.77
WRITING	9	.78	.75	.75	.77	.87	.83	.77
TOTAL SCALE	36	.92	.91	.90	.91	.96	.93	.91
GOAL - 8K	30	.63	.56	.61	.75	N.A.	N.A.	N.A.
VALUING	24	.66	.70	.72	.73	.90	.87	.75
RECEIVING	24	.82	.85	.86	.84	.93	.96	.86
TOTAL SCALE	48	.83	.86	.88	.86	.93	.96	.86
EFFSOLN	13	.71	.71	.78	.73	.93	.77	.77
INEFFSOL	13	.79	.83	.84	.82	.96	.82	.84
EMOTADJ	9	.70	.73	.79	.74	.69	.95	.73
TOTAL SCALE	35	.80	.83	.86	.83	.96	.86	.86

N.A. - Not Appropriate

RESPONSE CONSISTENCY OVER TIME

Two questions relate to stability of test scores. Do students stay in the same rank order position relative to one another from one test occasion to the next? Do students who are classified as having positive or negative attitudes remain in their respective categories across testing occasions? The first question is answered by computing correlations between the two test scores obtained at different points in time. The latter is answered by adding the percentage of students who stay in the favorable attitude classification with the percentage remaining in the unfavorable classification.

In February and March 1974, data on test stability for eight of the 11 scales over a four and one-half week period were obtained from 400 8th grade students in a large school district in southeastern Pennsylvania. Test administration followed procedures recommended in the EQA statewide *Monitor's Handbook*. A second study involving 120 students attending a school district in western Pennsylvania produced stability estimates for the remaining three scales.

Table 2 presents test-retest correlations developed from norm referenced scores on all sub and total scales. An index of continuity for scores on each of three criterion scoring levels is also given. This index tells us the per cent of students remaining in the same scoring category (i.e., those passing on the first occasion who also pass on the second occasion or those failing on the first occasion, who also fail on the second occasion). From this table it can be seen that the total scale reliabilities are quite high and range from .73 through .81. The continuity index figures show that a large percentage of students don't change their answers sufficiently across testing occassions to be placed in a different category.

TABLE 2

TEST-RETEST RELIABILITY AND CONTINUITY INDICES FOR SUB AND TOTAL EQA INSTRUMENTS

Scale Name	Test-Retest Reliability	Indices of Continuity (Criterion Scoring)		
		Level One	Level Two	Level Three
CSELF	.73	91%	81%	80%
CONTENV	.63	95%	87%	81%
RELATE	.79	94%	86%	85%
SCHLIMAG	.76	86%	79%	80%
TOTAL SCALE	.80	96%	88%	84%
RACE	.67	83%	78%	78%
RELIB	.72	82%	77%	79%
SES	.62	88%	80%	83%
INTELL	.64	80%	70%	80%
HANDCP	.69	86%	92%	73%
TOTAL SCALE	.76	91%	82%	83%
GOAL - 3V	.79	N.A.	N.A.	N.A.
GOAL - 3M	. 81	N.A.	N.A.	N.A.
LEARN	. 74	87%	80%	81%
SCHOOL	. 81	86%	84%	81%
TOTAL SCALE	. 81	88%	84%	81%
WELFDIG	. 67	71%	78%	85%
LAWAUTH	. 70	73%	79%	89%
RESPINTG	. 68	73%	79%	88%
TOTAL SCALE	. 73	77%	84%	91%
PHEALTH	. 67	71%	78%	85%
SAFETY	. 69	72%	77%	87%
DRUGS	. 75	87%	83%	81%
TOTAL SCALE	. 71	84%	76%	88%
VISLART	. 63	79%	75%	80%
PERFMART	. 54	73%	73%	80%
SCIENCE	. 60	84%	78%	76%
WRITING	. 56	84%	80%	76%
TOTAL SCALE	. 64	87%	80%	80%
GOAL - 8K	. 76	N.A.	N.A.	N.A.
VALUING	. 66	N.A.	N.A.	N.A.
RECEIVING	.75	N.A.	N.A.	N.A.
TOTAL SCALE	. 75	N.A.	N.A.	N.A.
EFFSOLN	. 63	87%	77%	75%
INEFFSOL	71	92%	84%	78%
EMOTADJ	68	70%	86%	87%
TOTAL SCALE	.9	91%	82%	77%

SECTION THREE

Validity of Goal Instruments

WHAT IS VALIDITY?

Validity is evidence–evidence that helps us separate fact from fiction concerning test results. In the case of the Educational Quality Assessment Program, information on the validity of each scale can help state educators translate the *paper goals* and *number outcomes* into a better understanding of Pennsylvania school age children. No single procedure or experimental design gives a complete picture of a test's validity. Instead each study, in its own way, helps round out a picture of what each instrument measures and, therefore, aids our interpretation of test outcomes.

Most test manuals at some point address themselves to questions of validity and reliability. Underpinning these concepts are the more basic questions of the interpretability and usefulness of information engendered from the test. The internal structure characteristics of the EQAI instruments which are presented in Section 3 offer initial evidence of the tests' useability by demonstrating empirically that the scales can elicit consistent student responses both within sets of similar items and across time. Low susceptibility to faking and ease of readability also support the premise that the vast majority of 8th grade students can interact easily with the battery in the test-setting.

The internal structure of tests also can be examined by asking students to describe their general reactions to each scale and asking professional educators to judge the appropriateness of contents of each scale.

ALLOWING STUDENTS TO REACT TO THE INVENTORY

During fieldtesting, a six-statement questionnaire was inserted at the end of each proposed EQAI scale. Students who were selected to respond to the questionnaire represented a range of high, average and low socioeconomic backgrounds attending urban, suburban and rural schools. Approximately 300 students per EQAI scale responded to the special questionnaire. For example, students who had completed the Goal I–Self-Esteem scale were asked:

1. *In general, do you feel that the questions get at self-understanding?*
   (Please circle Yes or No)

2. *Do you think you answered these questions honestly?*
   (Please circle Yes or No)

3. *Please write a sample question that would better reflect self-understanding.*

4. *Go back through the scale and identify those statements which you had difficulty understanding. Please write down the question numbers.*

5. *Please list the words which you had difficulty understanding.*

6. *General comments. Here is a chance for you to write any comments you might have about this scale.*

Data obtained from this procedure were used to refine the instruments and to obtain estimates of face validity. The percentage of respondents agreeing that particular tests appeared to reflect the trait of interest ranged from 93 per cent for the creative activities scale through 98 per cent for the Goal III mathematics test. An average 96 per cent felt that they had responded honestly to test items. Responses to question six indicate that fewer than 5 per cent of the students felt that time spent completing the scales was wasted (i.e. they thought the tests were *stupid*, not relevant, or unrealistic to their personal lives.)

JUDGMENTS OF TEST CONTENT BY EDUCATORS

When constructing tests to be used in a statewide assessment program, it is extremely important to make systematic efforts to insure that item and scale content are both logically and empirically related to the concepts they are designed to measure. From the outset, content specifications were developed for each goal area. Items and test formats were then designed to reflect these specifications as closely as possible. Working papers for each instrument–describing its item development and selection procedures–together with its *content map*, are available at the office of the Division of Educational Quality Assessment in Harrisburg.

During all stages of test development, curriculum specialists from the Department of Education, together with measurement researchers and local school district personnel, judged item content, scope and appropriateness. In addition, 40 teachers and administrators from the Carlisle Area School District who had undergone a 15-hour training program in quality assessment procedures were asked to rate each final instrument in terms of item and content appropriateness. The lowest agreement of content appropriateness was found for the creative activities scale (90 per cent thinking the test adequately measured the concept). Between 94 and 98 per cent agreement on content appropriateness was obtained for the other 10 scales.

STUDENT SCORES vs TEACHER PERCEPTIONS

The major purpose of the quality assessment procedure is to identify student progress in the 10 quality education goal areas. This will enable educators to more clearly focus curriculum efforts on student strenghths and weaknesses. The ability of teachers to identify student needs by classroom observation techniques is germane to meeting student needs or maintaining already high levels of student achievement. An important question in this regard is: *Do student test scores correspond to teacher judgments based upon observations of student classroom behaviors?*

Faculty and administrators from both the middle (N=8) and senior high schools (N=20) of Carlisle Area School District participated in a 15 hour in-service workshop sponsored jointly by the state educational department's Bureau of Curriculum Services, Capitol Area Intermediate Unit and the Carlisle Area School District. Training was conducted by Division of Quality Assessment personnel and was designed to communicate the nature and substance of the EQA program and to familiarize teachers with goal measurement rationales and test content. Concurrently the entire student body at both grades 8 and 11 completed appropriate forms of the EQA Inventory. After initial training in classroom observation techniques, teachers were asked to nominate the highest 20 per cent and lowest 20 per cent of their students in the areas of self-esteem, tolerance, attitude toward school and learning, citizenship, health habits, creative performance and coping with change.

Each student was then assigned a score based on the number of high or low nominations the student received. Each score was a ratio formed by taking the number of positive *high* nominations minus the number of negative *low* nominations divided by the number of possible nominations. Distributions formed from these scores were then examined in order to place students into high and low groups for the purpose of analysis.

The number of students in each group varied from goal to goal and typically ranged from 75 to 100. Differences in mean scores between the high/low groups were statistically evaluated via a series of Fisher-Behrens t tests which correct for biases associated with unequal variances when sample sizes are unequal.

Table 3 presents the mean scores for the high and low teacher nominated groups, the t value along with its associated level of statistical significance and the point biserial correlation coefficient between group membership and scale scores. Scale names are given in acronym form. As can be seen from the probability column, student scores on five of the seven composite instruments are significantly related to teacher perceptions.

Statistically significant relationships between teacher perception of classroom behavior and test scores are found for the total scale and each subscale for Goal I-self-esteem, Goal IV-interest in school, Goal V-citizenship and Goal VI-health habits. This indicates that teachers were readily able to observe characteristics in their students which were also picked up by the student's responses to the test items. The relationship found for Goal II is significant for the total scale and three of the five subscales. The race and religion subscales failed to reveal significant differences. Student tolerance toward others of a different race or religion may rarely be expressed in this school and, therefore, be observed by teachers too infrequently to serve as a reliable basis for classification. Only one subscale for Goal X-coping with change was significant. Again, circumstances in which teachers might observe students with regard to this characteristic may appear very seldom in the classroom environment. None of the Goal VII-creative performance scales showed reliable relationships with teacher perception.

Table 4 gives a criterion-referenced perspective to the results of the Carlisle study. The per cent of students passing at each of the criterion scoring levels for high and low groups are shown. Also given is the difference in the per cent passing between the groups and the level of statistical significance of this difference based on test results.

Note that the interest in school and citizenship instruments show significant differences across all three criterion levels.

The creative activities scale shows statistically reliable differences at the second criterion level. Recall from Table 3 that this scale failed to discriminate between groups when norm scoring procedures were used.

All seven instruments investigated in this study show the ability to discriminate between teacher-selected high and low groups on at least one, and in most cases more than one, criterion level.

The combined norm-referenced and criterion-referenced results give strong support to the notion that teachers, through close classroom observation, can identify generally the same students that the tests show to be above average or below average on the seven general traits measured by the EQAI.

TABLE 3

CORRESPONDENCE BETWEEN TEACHER NOMINATIONS AND TEST SCORES FOR SEVEN EQA INSTRUMENTS

SCALE NAME	Statistic				
	H Group	L Group	't'	Prob.	Pt. Biserial r
CSELF	19.27	16.77	2.48	.01	.23
CONTENV	21.07	19.21	2.36	.01	.22
RELATE	20.82	16.79	4.15	.0001	.37
SCHLIMAG	18.08	15.06	3.05	.001	.27
GOAL I TOTAL	79.25	67.98	3.68	.0001	.33
RACE	12.61	11.76	1.32	N.S.	.11
RELIG	12.08	10.98	1.45	N.S.	.12
SES	13.24	11.98	2.21	.05	.18
INTELL	11.73	10.61	2.08	.05	.17
HANDCP	13.33	12.32	1.82	.07	.14
GOAL II TOTAL	62.96	57.56	2.32	.05	.18
LEARN	44.07	38.62	3.90	.0001	.34
SCHOOL	42.58	32.00	6.06	.0001	.50
GOAL IV TOTAL	86.63	70.66	5.58	.0001	.47
WELFDIG	20.84	16.42	5.45	.0001	.37
LAWAUTH	26.91	20.23	5.94	.0001	.39
RESPINTG	22.46	17.77	4.53	.0001	.31
GOAL V TOTAL	70.21	54.52	6.13	.0001	.41
PHEALTH	29.80	26.92	2.55	.01	.19
SAFETY	19.32	17.21	2.14	.05	.16
DRUGS	25.39	21.20	4.76	.0001	.34
GOAL VI TOTAL	74.46	65.34	3.84	.0001	.28
VISLART	12.32	11.52	.67	N.S.	.05
PERFMART	8.76	8.25	.49	N.S.	.04
SCIENCE	13.33	13.23	.08	N.S.	.00
WRITING	11.43	13.36	- 1.70	N.S.	-.14
GOAL VII TOTAL	45.84	46.34	.12	N.S.	.01
EFFSOLN	24.22	23.33	1.12	N.S.	.09
INEFFSOL	28.21	25.49	2.51	.01	.19
EMOTADJ	12.79	14.13	1.66	N.S.	.14
GOAL X TOTAL	65.45	62.96	1.21	N.S.	.10

Note - N.S. - Not statistically significant at or above .05 probability level.

TABLE 4

CORRESPONDENCE OF STUDENT PLACEMENT INTO HIGH-LOW GROUPS BY TEACHER NOMINATION AND TEST SCORES FOR SEVEN EQA INSTRUMENTS: CRITERION-REFERENCED SCORING

Scale Name	Scoring Levels: Level One				Level Two				Level Three			
	% Passing Per Group				% Passing Per Group				% Passing Per Group			
	High	Low	Diff*	Prob.	High	Low	Diff*	Prob.	High	Low	Diff*	Prob.
Self Esteem	97	92	5	N.S.	89	61	28	.0001	69	43	27	.0020
Tolerance toward Others	91	85	6	N.S.	75	72	3	N.S.	40	26	14	.0050
Interest in School	96	72	24	.0004	80	45	35	.0001	67	42	35	.0001
Citizenship	76	30	36	.0001	39	13	26	.0001	18	6	12	.0100
Health Habits	96	72	24	.0001	60	43	17	.0100	16	11	5	N.S.
Creative Activities	88	79	9	N.S.	77	62	16	.0200	37	30	7	N.S.
Preparing for Change	99	96	3	N.S.	84	82	2	N.S.	23	10	13	.0300

Note: Diff* stands for difference between per cent passing in High vs Low groups.

RELATED STUDIES

The Division of Research and the Division of Educational Quality Assessment have sponsored a variety of independent studies designed to give further insight into the validity of the grade 8 battery. The abstracts associated with these studies are given on the following pages in goal number order. Complete data for these studies are available upon request at the office of the Division of Research. All abstracts have been prepared by the researcher responsible for conducting each particular study.

DO EQA GOAL II TEST SCORES AGREE WITH PEER RATINGS?

Gregory A. Shannon

The purpose of this study was to provide behavior-related criterion validity for the EQA Goal II tests at grade 8. This study investigated the extent to which the tolerant behaviors of the students related to their Goal II test scores. The behavioral criteria were peer ratings. The students were asked to select five of their classmates who would *most likely*, and five of their classmates who would *least likely*, perform each of four tolerant behaviors. The students were asked to indicate whether each selection was based upon race, religion, physical handicap, socioeconomic status, or intelligence.

The sample consisted of 62 white and 29 nonwhite 8th grade students from two urban Pennsylvania school districts for a total of 91 students. Three classrooms were selected from one district and one classroom from the other district.

From the total ratings observed for each student, high and low groups were selected and compared on the Goal II tests. In addition, the total ratings were correlated with the Goal II tests. Modest validity support was found for the religion, socioeconomic status, and intelligence subtests, and the total test. No support was found for either the race or the physical handicap subtest.

VALIDATION OF THE PENNSYLVANIA EDUCATIONAL QUALITY ASSESSMENT INSTRUMENTS TO MEASURE STUDENTS' ACCEPTANCE OF DIFFERING OTHERS

Peggy Stank

A sociometric measure was used to validate the Pennsylvania Educational Quality Assessment Goal II instruments for 5th, 8th and 11th grades. These instruments were developed to measure students' acceptance of persons who differ from them in race, socioeconomic status (SES), intelligence, religion and persons who have some handicap.

Both the sociometric measure and the Goal II instruments were administered to one 5th, one 8th and one 11th grade class in a Pennsylvania school district. All classes had adequate proportions of race, SES levels and intelligence levels for comparison of student scores on the two measures. Although it was not practical to compare EQA subscales on religion and handicapped with the sociometric measure, it seems reasonable to assume that these scales will perform similarly to SES and intelligence level difference subscales.

The basic assumption of this study was that if student scores on the EQA measures accurately predicted student behaviors toward differing others, then the EQA instruments could be accepted as having adequate validity for inclusion in the battery of tests. The results of the study showed that, with the exception of the race subscale in the 8th grade and the 11th grade subscale on intelligence levels, the EQA Goal II instruments accurately predicted student behavior as indicated by the sociometric measure.

WHICH STUDENT BEHAVIORS RELATE TO EQA GOAL IV TEST SCORES?

James R. Masters, Gregory A. Shannon and Francis J. Reardon

Students in three school districts representing a rural-suburban-urban continuum took the EQA Goal IV instrument. Additionally, a survey instrument which asked the students a number of questions about themselves was administered. High and low scoring students on the EQA instrument were then compared on variables that were obtained by the survey questionnaire.

Very large differences appeared between the two groups on such variables as how far the student wished to go in school, amount of time spent studying, occurrence of unexcused absence from school, grade point average and paying attention in class. Smaller, but still significant, differences existed in such variables as number of books read over a recent period and whether or not the student generally completed assignments on time. Data in the form of student conduct ratings were collected from an administrator at each school. The high and low scoring groups failed to differ on these measures.

Extremely high and low scores on the EQA Goal IV test differ markedly on several behavioral variables that indicate positive attitudes. Such a finding lends a great deal of support to the validity of the 8th grade instrument as a measure of student attitude toward school.

DISCRIMINATION OF CITIZENSHIP CHARACTERISTICS IN AN INSTITUTION FOR NEGLECTED AND DELINQUENT GIRLS

N. F. Russell

This study was conducted in a home for wayward girls located in western Pennsylvania. The institution has a *stated* policy of periodic appraisal (every three months) of each resident's social and emotional adjustment. Evaluations are made by a three-member team including a psychologist, a case worker and a cottage attendant. Characteristics which are evaluated include personal responsibility, honesty, fighting behaviors and ability to get along with others. Assignments to living quarters are made contingent upon the above evaluations. As a girl's adjustment improves, she is moved to another cottage with girls of similar adjustment and with more privileges. There are five cottage units in the institution with cottage number five housing the *most adjusted* of the students.

The scale was administered to the entire population of girls (N=46) living in the institution. For purposes of analysis the girls were separated into two groups. Those living in the first three cottages formed group one (*most maladjusted*, N=25) Group two (*most adjusted*, N=21) was composed of girls housed in cottages four and five.

Analyses revealed that the citizenship scale could (1) discriminate between these two groups (the adjusted group scoring higher), and (2) discriminate between the total institutional group and a group of public school children having similar home backgrounds (the public school group scoring higher).

THE RELATIONSHIP BETWEEN THE EQA GOAL VII SCORES AND SELECTED AREAS OF CREATIVE ATTITUDE AND OUTPUT

Gregory A. Shannon and James R. Masters

This study was designed to obtain validity support for the EQA Goal VII subtests and the total test when scored by either the norm-referenced or the criterion-referenced method. The sample consisted of 38 grade 8 students from a suburban Pennsylvania school district.

The criterion measures included a questionnaire and a semantic differential instrument. For each activity described in the Goal VII test, the questionnaire asked about the students' willingness to do the activity, how much they felt their school would encourage them in it, the grade level at which the students last performed the activity, and the number of times each activity was performed during the previous two years. The second instrument asked the students to respond to adjective word-pairs describing how they felt about working on visual arts, performing arts, writing and science projects. This instrument yielded measures of how much the students would like to do the activities, how competent they felt about doing the activities, and the extent to which they felt that their teacher would encourage them.

Validity support was found for the total Goal VII test and all of the subtests when scored by either of the two scoring methods, except the output scores of the visual arts subtest. Students who earned high scores on the total test tended to enjoy working on creative activities. Students who earned high scores on the subtests tended to have a positive attitude toward doing the activities described by each subtest. Students who earned high scores on the performing arts, writing and science subtests tended to have performed these activities at either the 7th or the 8th grade level.

Thus, the studies done in 1974 support the validity of this test by showing a positive relationship between EQA Goal VII test scores and self-reports of student behaviors and attitudes.

GOAL IX-GRADE 8 VALIDITY STUDY

Grace E. Laverty and Joyce S. Kim

Goal IX was validated by a correlational study with scores from the EQA inventory subscales compared with choices made in the same content area on a survey of possible field trips and presentations.

The sample included 66 eighth grade pupils from a school representing suburban and rural communities and various socioeconomic levels. The 8th grade sample was taken from two social studies classes which were completely heterogeneous groups. Thus, the sample for testing was fairly representative of the 8th graders in the school.

The pupils completed the EQA inventory and a survey of special events presented as possible field trips or special programs. They were asked to select as many of the 18 events as they would be interested in attending. The analysis compared their choices in each of the nine content areas with their score on the inventory items representing that content area.

A similar study was done at grades 5, 8 and 11. Over all grades and content areas, representing 30 separate correlations, coefficients were positive except for one. The 8th grade analysis showed positive correlations in all areas and on the overall test, and all were statistically significant except world events and science. In general, the evidence gathered by comparing actual choices by students in the survey of special events with their scores in the related area of Goal IX Inventory--suggested that the instrument is measuring attitudes and values in these areas of human accomplishment.

DO EQA GOAL X SCORES AGREE WITH TEACHER RATINGS OF STUDENTS' ABILITY TO ADJUST TO CHANGE?

James R. Masters and Gregory A. Shannon

A study of the instrument's validity was conducted in a suburban school district where a large percentage of the student population had undergone change in their lives. Approximately 60 8th grade students who had experienced a great deal of change (termed the Change group), and 60 8th grade students who had experienced little change (termed the Nonchange group) participated in the study. The Change group had lived through such upsets as parental divorces, separations or deaths, or had experienced at least two residential movements. Because they had lived through change, their day-to-day behaviors would reflect how well they had been able to deal with it. Also, since situations described in the EQA instrument would be similar to those they had experienced, their predictions of behavior would be expected to be more accurate than those of others in their schools.

From the 120 students, teachers were asked to choose students *high* and *low* in *emotional fortitude* --defined as the ability to 1) recover from a serious emotional setback, 2) confront difficult obstacles, and 3) discipline and direct one's own behavior in order to achieve a goal.

For the Change group, those rated *high* scored higher than those rated *low* on both the ineffective solutions subscale and the total instrument. For the Nonchange group no differences were found between *highs* and *lows* for any subscale or for the total instrument.

Support was found, therefore, for the validity of both the ineffective solutions subscale and the total instrument. In addition, students had little difficulty placing themselves in the situations presented in the instrument, and 92 per cent felt confident that their responses accurately reflected their actual behaviors.

VALIDITY INFORMATION FROM FACTOR ANALYSIS

Factor analysis is a term used to describe a set of statistical procedures which can be used to analyze the intercorrelations among a set of variables such as test scores. Factor analysis aids us in interpreting these relationships in terms of underlying *factors* and gives us insight into the amount of variation in each separate EQAI instrument which is associated with each of the hypothetical factors. In the general sense, factor analysis presents a picture of how students respond to the test battery in its entirety by showing how the scale scores within the battery cluster together.

Factor analysis begins with a correlation matrix of all the instrument scores contained in the battery. For the purposes of this publication, the matrix presented in Table 5 is limited to the 26 subscales and the math and verbal instruments. All subscale names are given in acronym form due to space limitations. This table has two X and Y columns per page. In each column set acronyms for two subscales are given together with the correlation coefficient for that pair. Correlations are based upon approximately 3500 student records selected from the March 1974 testing period which constitute a systematic sample from the data tape containing about 50,000 student records.

A series of analyses* were performed on this correlational data which helped identify the underlying factor structure of the EQA Inventory. This structure may be inferred by examining the clusters of subscales displaying correlations to a given factor. Results of these analyses are presented in Table 6.

This table lists the subscales in the left hand column. The coefficients in the body of the table are correlations between each subscale and the eight factors listed horizontally near the top of the table. These correlations are called *factor loadings*, and define the factor by showing how much each subscale is related to it. The highest loadings for each factor are highlighted by a box drawn around them. By looking to the subscale names associated with these loadings one can understand which subscales describe the factor.

FACTOR ONE: RESPONSIBILITY FOR WELL-BEING OF SELF AND OTHERS

Factor one appears to be composed of the citizenship (Goal V) and health (Goal VI) subscales. That these two goal areas cluster is due in part to methods used to obtain scores. Both tests have identical formats: persons are asked to put themselves in a *make-believe* student's place and decide whether to take *ideal* actions under a variety of motivation-including conditions, (i.e., peer pressure, reward, etc).

Although the clustering of these subscales appears to be artifact of type of format used to elicit students' responses, a common variable also seems to be mirrored by this factor. Each of the subscales in this cluster require that the students exhibit a sense of personal responsibility for their own well-being and the well-being of others in relation to health, safety and social interactions.

---

* Extraction of principal components followed by variance rotation–See Harmen, H. H., *Modern Factor Analysis*, Chicago Press, 1960.

FACTOR TWO: ATTITUDE TOWARD CREATIVE ENDEAVORS

This factor is chiefly defined by the creative activities subscales indicating that the Goal VII instrument tends to give information about students, which is relatively independent from that offered by the other EQAI scales. The relatively high correlation of this factor to the Goal IX receiving subscale, shows that people who like to participate in science, art, writing and performing activities also like to observe others who are recognized as being proficient in similar areas.

FACTOR THREE: SELF-ESTEEM

The subscale cluster comprising this factor is basically associated with the self-esteem instrument. In addition, a clear relationship is also found between this factor and the interest in school (Goal IV) subscale. This is understandable since many of the items in the Goal I test attempt to assess self-esteem in the context of the classroom environment. Therefore, it is not surprising to find that some students who have a good self-concept in the school setting also tend generally to have positive feelings about their school experiences.

FACTOR FOUR: TOLERANCE TOWARD OTHERS

How comfortable students feel when coming into contact with others differing in racial, religious, wealth, intellectual, or physical attributes forms the base of this factor. This feeling state also has an action counterpart which is mirrored by the tendency to refrain from behaviors that might harm others. Evidence for this action component is shown by the significant positive relationship between the Goal V citizenship subscale-concern for welfare and dignity of others and this factor. The Goal IX valuing human accomplishments subscale also relates to this factor, demonstrating that the tendency to stereotype others in a negative way is also related to the person's intolerance of others.

FACTOR FIVE: BASIC COGNITIVE SKILLS

This factor is composed of the three basic skills measures (math, verbal and vocational knowledge) and the Goal VI drugs subscale. The positive correlation between the Goal I control of environment subscale and this factor suggests that students who are more successful in school achievement believe that they can influence, to a greater extent, their personal destinies.

FACTOR SIX: VALUING THE EDUCATIONAL EXPERIENCE

This factor is complex in that it is defined by various subscales across four goal areas. Included in this factor are the Goal I image in school settings, Goal IV attitude toward school and interest in learning, Goal IX valuing and receiving and Goal X effective solution subscales. All of these scales relate to attitudes and behaviors associated with classroom or school-related settings.

It appears from the way these subscales cluster that this factor might represent a belief that the classroom, school and general community are all fertile grounds to engage in learning experiences.

FACTOR SEVEN: EMOTIONAL ADJUSTMENT TO CHANGE

This factor is defined by two subscales in the Goal X coping with change instrument. The length of time necessary to adjust emotionally to frustration is seen to be strongly associated with the tendency to refrain from aggressive or withdrawal reactions to frustrating events. A

vicious circle might be reflected here in that people who get very upset in the face of change might be more willing to try negative solutions to their personal problems, leading to a new problem and, hence, a continuation of anxiety.

FACTOR EIGHT: PERSONAL AND EMOTIONAL HEALTH

This factor is basically defined by the Goal VI health, drug and safety subscales together with the effective and ineffective solution subscales of Goal X. All five of these subscales relate to student willingness to take various actions in social settings which relate to their physical or mental well-being. Interpretation of this factor suggests that people who are less able to cope with frustration tend to take greater risks in areas of safety, health and drug areas.

Tables 7 and 8 present additional statistics resulting from factor analytic procedures. The proportion of variation within each subscale that can be accounted for by the eight factor solution is shown in Table 7. The *explained* variation ranges from .54 for the religion subscale to a high of .87 for the emotional adjustment subscale. The Percent of Trace column in Table 8 shows what proportion of variability in student scores in the entire test battery can be explained by each of the eight factors.

TABLE 5

INDIVIDUAL STUDENT DATA *** GRADE 8 *** <<1974>>

X	PAIR	Y	R	X	PAIR	Y	R
CONTENV	2-	1 GSELF	0.503	GOAL-3M	11-	6 RELIG	0.139
RELATE	3-	1 GSELF	0.600	GOAL-3M	11-	7 SES	0.087
RELATE	3-	2 CONTENV	0.385	GOAL-3M	11-	8 INTELL	0.004
SCHLIMAG	4-	1 GSELF	0.557	GOAL-3M	11-	9 HANDCP	0.055
SCHLIMAG	4-	2 CONTENV	0.471	GOAL-3M	11-	10 GOAL-3V	0.692
SCHLIMAG	4-	3 RELATE	0.406	LEARN	12-	1 GSELF	0.193
RACEG2	5-	1 GSELF	0.031	LEARN	12-	2 CONTENV	0.352
RACEG2	5-	2 CONTENV	0.179	LEARN	12-	3 RELATE	0.134
RACEG2	5-	3 RELATE	0.116	LEARN	12-	4 SCHLIMAG	0.362
RACEG2	5-	4 SCHLIMAG	0.132	LEARN	12-	5 RACEG2	0.169
RELIG	6-	1 GSELF	0.077	LEARN	12-	6 RELIG	0.113
RELIG	6-	2 CONTENV	0.158	LEARN	12-	7 SES	0.166
RELIG	6-	3 RELATE	0.119	LEARN	12-	8 INTELL	0.200
RELIG	6-	4 SCHLIMAG	0.160	LEARN	12-	9 HANDCP	0.283
RELIG	6-	5 RACEG2	0.473	LEARN	12-	10 GOAL-3V	0.158
SES	7-	1 GSELF	0.102	LEARN	12-	11 GOAL-3M	0.164
SES	7-	2 CONTENV	0.189	SCHOOL	13-	1 GSELF	0.238
SES	7-	3 RELATE	0.115	SCHOOL	13-	2 CONTENV	0.397
SES	7-	4 SCHLIMAG	0.137	SCHOOL	13-	3 RELATE	0.160
SES	7-	5 RACEG2	0.508	SCHOOL	13-	4 SCHLIMAG	0.453
SES	7-	6 RELIG	0.452	SCHOOL	13-	5 RACEG2	0.181
INTELL	8-	1 GSELF	0.038	SCHOOL	13-	6 RELIG	0.117
INTELL	8-	2 CONTENV	0.146	SCHOOL	13-	7 SES	0.143
INTELL	8-	3 RELATE	0.057	SCHOOL	13-	8 INTELL	0.186
INTELL	8-	4 SCHLIMAG	0.108	SCHOOL	13-	9 HANDCP	0.218
INTELL	8-	5 RACEG2	0.434	SCHOOL	13-	10 GOAL-3V	0.221
INTELL	8-	6 RELIG	0.325	SCHOOL	13-	11 GOAL-3M	0.225
INTELL	8-	7 SES	0.502	SCHOOL	13-	12 LEARN	0.691
HANDCP	9-	1 GSELF	0.098	WELFDIG	14-	1 GSELF	0.099
HANDCP	9-	2 CONTENV	0.233	WELFDIG	14-	2 CONTENV	0.272
HANDCP	9-	3 RELATE	0.124	WELFDIG	14-	3 RELATE	0.070
HANDCP	9-	4 SCHLIMAG	0.148	WELFDIG	14-	4 SCHLIMAG	0.216
HANDCP	9-	5 RACEG2	0.484	WELFDIG	14-	5 RACEG2	0.307
HANDCP	9-	6 RELIG	0.327	WELFDIG	14-	6 RELIG	0.172
HANDCP	9-	7 SES	0.520	WELFDIG	14-	7 SES	0.250
HANDCP	9-	8 INTELL	0.574	WELFDIG	14-	8 INTELL	0.292
GOAL-3V	10-	1 GSELF	0.114	WELFDIG	14-	9 HANDCP	0.339
GOAL-3V	10-	2 CONTENV	0.299	WELFDIG	14-	10 GOAL-3V	0.202
GOAL-3V	10-	3 RELATE	0.091	WELFDIG	14-	11 GOAL-3M	0.175
GOAL-3V	10-	4 SCHLIMAG	0.233	WELFDIG	14-	12 LEARN	0.321
GOAL-3V	10-	5 RACEG2	0.180	WELFDIG	14-	13 SCHOOL	0.344
GOAL-3V	10-	6 RELIG	0.200	LAWAUTH	15-	1 GSELF	0.102
GOAL-3V	10-	7 SES	0.120	LAWAUTH	15-	2 CONTENV	0.234
GOAL-3V	10-	8 INTELL	0.035	LAWAUTH	15-	3 RELATE	-0.013
GOAL-3V	10-	9 HANDCP	0.118	LAWAUTH	15-	4 SCHLIMAG	0.233
GOAL-3M	11-	1 GSELF	0.102	LAWAUTH	15-	5 RACEG2	0.174
GOAL-3M	11-	2 CONTENV	0.291	LAWAUTH	15-	6 RELIG	0.057
GOAL-3M	11-	3 RELATE	0.085	LAWAUTH	15-	7 SES	0.147
GOAL-3M	11-	4 SCHLIMAG	0.233	LAWAUTH	15-	8 INTELL	0.244
GOAL-3M	11-	5 RACEG2	0.122	LAWAUTH	15-	9 HANDCP	0.239

INDIVIDUAL STUDENT DATA *** GRADE 8 *** <<1974>>

PAIR X			Y	R	PAIR X			Y	R
LAWAUTH	15-	10	GOAL-3V	0.081	SAFETY	18-	15	LAWAUTH	0.548
LAWAUTH	15-	11	GOAL-3M	0.074	SAFETY	18-	16	RESPINTG	0.409
LAWAUTH	15-	12	LEARN	0.353	SAFETY	18-	17	PHEALTH	0.499
LAWAUTH	15-	13	SCHOOL	0.362	DRUGS	19-	1	GSELF	0.209
LAWAUTH	15-	14	WELFDIG	0.630	DRUGS	19-	2	CONTENV	0.337
RESPINTG	16-	1	GSELF	0.048	DRUGS	19-	3	RELATE	0.041
RESPINTG	16-	2	CONTENV	0.226	DRUGS	19-	4	SCHLIMAG	0.268
RESPINTG	16-	3	RELATE	0.030	DRUGS	19-	5	RACEG2	0.118
RESPINTG	16-	4	SCHLIMAG	0.215	DRUGS	19-	6	RELIG	0.038
RESPINTG	16-	5	RACEG2	0.263	DRUGS	19-	7	SES	0.132
RESPINTG	16-	6	RELIG	0.152	DRUGS	19-	8	INTELL	0.134
RESPINTG	16-	7	SES	0.235	DRUGS	19-	9	HANDCP	0.194
RESPINTG	16-	8	INTELL	0.298	DRUGS	19-	10	GOAL-3V	0.323
RESPINTG	16-	9	HANDCP	0.295	DRUGS	19-	11	GOAL-3M	0.313
RESPINTG	16-	10	GOAL-3V	0.101	DRUGS	19-	12	LEARN	0.347
RESPINTG	16-	11	GOAL-3M	0.083	DRUGS	19-	13	SCHOOL	0.393
RESPINTG	16-	12	LEARN	0.353	DRUGS	19-	14	WELFDIG	0.391
RESPINTG	16-	13	SCHOOL	0.382	DRUGS	19-	15	LAWAUTH	0.472
RESPINTG	16-	14	WELFDIG	0.638	DRUGS	19-	16	RESPINTG	0.343
RESPINTG	16-	15	LAWAUTH	0.671	DRUGS	19-	17	PHEALTH	0.478
PHEALTH	17-	1	GSELF	0.150	DRUGS	19-	18	SAFETY	0.572
PHEALTH	17-	2	CONTENV	0.247	VISLART	20-	1	GSELF	0.043
PHEALTH	17-	3	RELATE	0.028	VISLART	20-	2	CONTENV	0.039
PHEALTH	17-	4	SCHLIMAG	0.204	VISLART	20-	3	RELATE	0.106
PHEALTH	17-	5	RACEG2	0.097	VISLART	20-	4	SCHLIMAG	0.136
PHEALTH	17-	6	RELIG	0.053	VISLART	20-	5	RACEG2	0.114
PHEALTH	17-	7	SES	0.102	VISLART	20-	6	RELIG	0.101
PHEALTH	17-	8	INTELL	0.151	VISLART	20-	7	SES	0.061
PHEALTH	17-	9	HANDCP	0.193	VISLART	20-	8	INTELL	0.099
PHEALTH	17-	10	GOAL-3V	0.154	VISLART	20-	9	HANDCP	0.116
PHEALTH	17-	11	GOAL-3M	0.122	VISLART	20-	10	GOAL-3V	-0.059
PHEALTH	17-	12	LEARN	0.320	VISLART	20-	11	GOAL-3M	-0.065
PHEALTH	17-	13	SCHOOL	0.279	VISLART	20-	12	LEARN	0.175
PHEALTH	17-	14	WELFDIG	0.476	VISLART	20-	13	SCHOOL	0.120
PHEALTH	17-	15	LAWAUTH	0.531	VISLART	20-	14	WELFDIG	0.085
PHEALTH	17-	16	RESPINTG	0.435	VISLART	20-	15	LAWAUTH	0.089
SAFETY	18-	1	GSELF	0.121	VISLART	20-	16	RESPINTG	0.119
SAFETY	18-	2	CONTENV	0.278	VISLART	20-	17	PHEALTH	0.029
SAFETY	18-	3	RELATE	0.0[illegible]	VISLART	20-	18	SAFETY	0.025
SAFETY	18-	4	SCHLIMAG	0.192	VISLART	20-	19	DRUGS	-0.049
SAFETY	18-	5	RACEG2	0.137	PERFMART	21-	1	GSELF	0.032
SAFETY	18-	6	RELIG	0.041	PERFMART	21-	2	CONTENV	-0.044
SAFETY	18-	7	SES	0.115	PERFMART	21-	3	RELATE	0.120
SAFETY	18-	8	INTELL	0.189	PERFMART	21-	4	SCHLIMAG	0.082
SAFETY	18-	9	HANDCP	0.225	PERFMART	21-	5	RACEG2	0.072
SAFETY	18-	10	GOAL-3V	0.161	PERFMART	21-	6	RELIG	0.084
SAFETY	18-	11	GOAL-3M	0.137	PERFMART	21-	7	SES	0.053
SAFETY	18-	12	LEARN	0.330	PERFMART	21-	8	INTELL	0.089
SAFETY	18-	13	SCHOOL	0.343	PERFMART	21-	9	HANDCP	0.083
SAFETY	18-	14	WELFDIG	0.434	PERFMART	21-	10	GOAL-3V	-0.142

INDIVIDUAL STUDENT DATA *** GRADE 8 *** <<1974>>

X	PAIR	Y	R	X	PAIR	Y	R
PERFMART	21- 11	GOAL-3M	-0.139	WRITING	23- 20	VISLART	0.560
PERFMART	21- 12	LEARN	0.099	WRITING	23- 21	PERFMART	0.614
PERFMART	21- 13	SCHOOL	0.021	WRITING	23- 22	SCIENCE	0.661
PERFMART	21- 14	WELFDIG	0.041	GOAL-8K	24- 1	GSELF	0.130
PERFMART	21- 15	LAWAUTH	0.044	GOAL-8K	24- 2	CONTENV	0.312
PERFMART	21- 16	RESPINTG	0.068	GOAL-8K	24- 3	RELATE	0.106
PERFMART	21- 17	PHEALTH	-0.004	GOAL-8K	24- 4	SCHLIMAG	0.208
PERFMART	21- 18	SAFETY	-0.028	GOAL-8K	24- 5	RACEG2	0.146
PERFMART	21- 19	DRUGS	-0.115	GOAL-8K	24- 6	RELIG	0.166
PERFMART	21- 20	VISLART	0.671	GOAL-8K	24- 7	SES	0.100
SCIENCE	22- 1	GSELF	0.037	GOAL-8K	24- 8	INTELL	0.036
SCIENCE	22- 2	CONTENV	-0.014	GOAL-8K	24- 9	HANDCP	0.131
SCIENCE	22- 3	RELATE	0.103	GOAL-8K	24- 10	GOAL-3V	0.643
SCIENCE	22- 4	SCHLIMAG	0.077	GOAL-8K	24- 11	GOAL-3M	0.581
SCIENCE	22- 5	RACEG2	0.088	GOAL-8K	24- 12	LEARN	0.153
SCIENCE	22- 6	RELIG	0.079	GOAL-8K	24- 13	SCHOOL	0.210
SCIENCE	22- 7	SES	0.092	GOAL-8K	24- 14	WELFDIG	0.182
SCIENCE	22- 8	INTELL	0.117	GOAL-8K	24- 15	LAWAUTH	0.072
SCIENCE	22- 9	HANDCP	0.120	GOAL-8K	24- 16	RESPINTG	0.089
SCIENCE	22- 10	GOAL-3V	-0.074	GOAL-8K	24- 17	PHEALTH	0.144
SCIENCE	22- 11	GOAL-3M	-0.086	GOAL-8K	24- 18	SAFETY	0.143
SCIENCE	22- 12	LEARN	0.124	GOAL-8K	24- 19	DRUGS	0.307
SCIENCE	22- 13	SCHOOL	0.061	GOAL-8K	24- 20	VISLART	-0.076
SCIENCE	22- 14	WELFDIG	0.054	GOAL-8K	24- 21	PERFMART	-0.162
SCIENCE	22- 15	LAWAUTH	0.061	GOAL-8K	24- 22	SCIENCE	-0.097
SCIENCE	22- 16	RESPINTG	0.080	GOAL-8K	24- 23	WRITING	-0.123
SCIENCE	22- 17	PHEALTH	0.029	VALUING	25- 1	GSELF	0.167
SCIENCE	22- 18	SAFETY	0.015	VALUING	25- 2	CONTENV	0.392
SCIENCE	22- 19	DRUGS	-0.050	VALUING	25- 3	RELATE	0.139
SCIENCE	22- 20	VISLART	0.647	VALUING	25- 4	SCHLIMAG	0.303
SCIENCE	22- 21	PERFMART	0.664	VALUING	25- 5	RACEG2	0.248
WRITING	23- 1	GSELF	0.070	VALUING	25- 6	RELIG	0.221
WRITING	23- 2	CONTENV	-0.050	VALUING	25- 7	SES	0.249
WRITING	23- 3	RELATE	0.072	VALUING	25- 8	INTELL	0.207
WRITING	23- 4	SCHLIMAG	0.098	VALUING	25- 9	HANDCP	0.285
WRITING	23- 5	RACEG2	0.027	VALUING	25- 10	GOAL-3V	0.358
WRITING	23- 6	RELIG	0.053	VALUING	25- 11	GOAL-3M	0.333
WRITING	23- 7	SES	0.024	VALUING	25- 12	LEARN	0.410
WRITING	23- 8	INTELL	0.076	VALUING	25- 13	SCHOOL	0.435
WRITING	23- 9	HANDCP	0.040	VALUING	25- 14	WELFDIG	0.341
WRITING	23- 10	GOAL-3V	-0.115	VALUING	25- 15	LAWAUTH	0.264
WRITING	23- 11	GOAL-3M	-0.114	VALUING	25- 16	RESPINTG	0.309
WRITING	23- 12	LEARN	0.109	VALUING	25- 17	PHEALTH	0.254
WRITING	23- 13	SCHOOL	0.021	VALUING	25- 18	SAFETY	0.281
WRITING	23- 14	WELFDIG	0.024	VALUING	25- 19	DRUGS	0.329
WRITING	23- 15	LAWAUTH	0.026	VALUING	25- 20	VISLART	0.052
WRITING	23- 16	RESPINTG	0.019	VALUING	25- 21	PERFMART	-0.031
WRITING	23- 17	PHEALTH	0.016	VALUING	25- 22	SCIENCE	0.016
WRITING	23- 18	SAFETY	-0.056	VALUING	25- 23	WRITING	-0.038
WRITING	23- 19	DRUGS	-0.055	VALUING	25- 24	GOAL-8K	0.369

INDIVIDUAL STUDENT DATA *** GRADE 8 *** <<1974>>

PAIR X		Y	R	PAIR X		Y	R
RECEIVNG	26- 1	GSELF	0.120	EFFSOLN	27- 26	RECEIVNG	0.319
RECEIVNG	26- 2	CONTENV	0.231	INEFFSOL	28- 1	GSELF	0.259
RECEIVNG	26- 3	RELATE	0.103	INEFFSOL	28- 2	CONTENV	0.397
RECEIVNG	26- 4	SCHLIMAG	0.270	INEFFSOL	28- 3	RELATE	0.129
RECEIVNG	26- 5	RACEG2	0.214	INEFFSOL	28- 4	SCHLIMAG	0.314
RECEIVNG	26- 6	RELIG	0.180	INEFFSOL	28- 5	RACEG2	0.182
RECEIVNG	26- 7	SES	0.231	INEFFSOL	28- 6	RELIG	0.128
RECEIVNG	26- 8	INTELL	0.278	INEFFSOL	28- 7	SES	0.216
RECEIVNG	26- 9	HANDCP	0.304	INEFFSOL	28- 8	INTELL	0.198
RECEIVNG	26- 10	GOAL-3V	0.102	INEFFSOL	28- 9	HANDCP	0.258
RECEIVNG	26- 11	GOAL-3M	0.089	INEFFSOL	28- 10	GOAL-3V	0.201
RECEIVNG	26- 12	LEARN	0.469	INEFFSOL	28- 11	GOAL-3M	0.171
RECEIVNG	26- 13	SCHOOL	0.413	INEFFSOL	28- 12	LEARN	0.374
RECEIVNG	26- 14	WELFDIG	0.319	INEFFSOL	28- 13	SCHOOL	0.446
RECEIVNG	26- 15	LAWAUTH	0.321	INEFFSOL	28- 14	WELFDIG	0.352
RECEIVNG	26- 16	RESPINTG	0.362	INEFFSOL	28- 15	LAWAUTH	0.384
RECEIVNG	26- 17	PHEALTH	0.248	INEFFSOL	28- 16	RESPINTG	0.379
RECEIVNG	26- 18	SAFETY	0.247	INEFFSOL	28- 17	PHEALTH	0.349
RECEIVNG	26- 19	DRUGS	0.213	INEFFSOL	28- 18	SAFETY	0.448
RECEIVNG	26- 20	VISLART	0.237	INEFFSOL	28- 19	DRUGS	0.446
RECEIVNG	26- 21	PERFMART	0.192	INEFFSOL	28- 20	VISLART	0.028
RECEIVNG	26- 22	SCIENCE	0.213	INEFFSOL	28- 21	PERFMART	-0.031
RECEIVNG	26- 23	WRITING	0.173	INEFFSOL	28- 22	SCIENCE	-0.001
RECEIVNG	26- 24	GOAL-8K	0.119	INEFFSOL	28- 23	WRITING	-0.058
RECEIVNG	26- 25	VALUING	0.513	INEFFSOL	28- 24	GOAL-8K	0.189
EFFSOLN	27- 1	GSELF	0.137	INEFFSOL	28- 25	VALUING	0.350
EFFSOLN	27- 2	CONTENV	0.253	INEFFSOL	28- 26	RECEIVNG	0.285
EFFSOLN	27- 3	RELATE	0.082	INEFFSOL	28- 27	EFFSOLN	0.366
EFFSOLN	27- 4	SCHLIMAG	0.249	EMOTADJ	29- 1	GSELF	0.155
EFFSOLN	27- 5	RACEG2	0.085	EMOTADJ	29- 2	CONTENV	0.136
EFFSOLN	27- 6	RELIG	0.024	EMOTADJ	29- 3	RELATE	0.064
EFFSOLN	27- 7	SES	0.117	EMOTADJ	29- 4	SCHLIMAG	0.117
EFFSOLN	27- 8	INTELL	0.174	EMOTADJ	29- 5	RACEG2	0.110
EFFSOLN	27- 9	HANDCP	0.197	EMOTADJ	29- 6	RELIG	0.104
EFFSOLN	27- 10	GOAL-3V	0.062	EMOTADJ	29- 7	SES	0.090
EFFSOLN	27- 11	GOAL-3M	0.060	EMOTADJ	29- 8	INTELL	0.084
EFFSOLN	27- 12	LEARN	0.454	EMOTADJ	29- 9	HANDCP	0.098
EFFSOLN	27- 13	SCHOOL	0.396	EMOTADJ	29- 10	GOAL-3V	0.056
EFFSOLN	27- 14	WELFDIG	0.276	EMOTADJ	29- 11	GOAL-3M	0.043
EFFSOLN	27- 15	LAWAUTH	0.370	EMOTADJ	29- 12	LEARN	0.067
EFFSOLN	27- 16	RESPINTG	0.327	EMOTADJ	29- 13	SCHOOL	0.097
EFFSOLN	27- 17	PHEALTH	0.314	EMOTADJ	29- 14	WELFDIG	0.097
EFFSOLN	27- 18	SAFETY	0.299	EMOTADJ	29- 15	LAWAUTH	0.080
EFFSOLN	27- 19	DRUGS	0.290	EMOTADJ	29- 16	RESPINTG	0.083
EFFSOLN	27- 20	VISLART	0.098	EMOTADJ	29- 17	PHEALTH	0.135
EFFSOLN	27- 21	PERFMART	0.034	EMOTADJ	29- 18	SAFETY	0.122
EFFSOLN	27- 22	SCIENCE	0.057	EMOTADJ	29- 19	DRUGS	0.121
EFFSOLN	27- 23	WRITING	0.039	EMOTADJ	29- 20	VISLART	-0.030
EFFSOLN	27- 24	GOAL-8K	0.101	EMOTADJ	29- 21	PERFMART	-0.025
EFFSOLN	27- 25	VALUING	0.267	EMOTADJ	29- 22	SCIENCE	-0.012

INDIVIDUAL STUDENT DATA *** GRADE 8 *** <<1974>>

PAIR X		Y	R
EMOTADJ	29- 23	WRITING	0.041
EMOTADJ	29- 24	GOAL-8K	0.059
EMOTADJ	29- 25	VALUING	0.051
EMOTADJ	29- 26	RECEIVNG	0.028
EMOTADJ	29- 27	EFFSOLN	-0.018
EMOTADJ	29- 28	INEFFSOL	0.406

TABLE 6

INDIVIDUAL STUDENT DATA *** GRADE 8 *** <<1974>>

THE ROTATED MATRIX OF FACTOR LOADINGS

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7	FACTOR 8
GSELF	1	0.02891	0.02208	0.87548	0.00854	0.02464	0.05452	-0.10014	-0.12853
CONTENV	2	0.14418	-0.06600	0.60631	0.13702	0.26419	0.27641	-0.09760	-0.15476
RELATE	3	-0.02930	0.08852	0.82198	0.10129	0.01558	0.00163	0.04519	0.07724
SCHLIMAG	4	0.14335	0.08324	0.66995	0.03123	0.16280	0.33203	-0.08081	0.01405
RACEG2	5	0.16005	0.04523	0.03667	0.73087	0.13584	0.06001	-0.06967	0.14501
RELIG	6	0.05600	0.06592	0.08988	0.61686	0.19998	0.03453	-0.12608	0.30731
SES	7	0.05398	0.01222	0.08263	0.79452	0.04491	0.07328	-0.03590	-0.03912
INTELL	8	0.12360	0.05914	-0.00691	0.75217	-0.08603	0.13325	0.01728	-0.18370
HANDCP	9	0.11567	0.05429	0.06312	0.75821	0.01150	0.17431	0.00665	-0.22070
GOAL-3V	10	0.06908	-0.08399	0.08197	0.08633	0.86949	0.06673	-0.02736	-0.03816
GOAL-3M	11	0.05213	-0.06783	0.08202	0.01448	0.84517	0.08889	-0.01040	-0.03111
LEARN	12	0.16233	0.10176	0.13278	0.09325	0.06760	0.77354	-0.05129	-0.17036
SCHOOL	13	0.20471	0.01149	0.20572	0.05095	0.14106	0.74620	-0.12266	-0.09628
WELFDIG	14	0.79551	0.02336	0.07125	0.23637	0.13124	0.15076	-0.01681	0.02780
LAWAUTH	15	0.82865	0.02970	0.03910	0.08037	-0.02083	0.21863	-0.02751	-0.15946
RESPINTG	16	0.79414	0.03152	0.01720	0.19489	-0.00939	0.27809	-0.03326	0.09235
PHEALTH	17	0.63626	0.01912	0.08493	0.02549	0.09590	0.11949	-0.09369	-0.36210
SAFETY	18	0.54430	-0.02072	0.03732	0.07733	0.10805	0.17357	-0.13672	-0.53434
DRUGS	19	0.42601	-0.07409	0.11669	0.04000	0.33169	0.18708	-0.12710	-0.56762
VISLART	20	0.04902	0.82731	0.04053	0.05725	-0.01985	0.12896	0.03232	0.04102
PERFMART	21	0.02671	0.85387	0.04092	0.05202	-0.11218	0.02373	0.02452	0.05044
SCIENCE	22	0.01194	0.86347	0.01376	0.08271	-0.03185	0.04202	0.01294	-0.03962
WRITING	23	-0.01943	0.83211	0.03294	0.00139	-0.06935	0.01632	-0.04471	-0.01731
GOAL-8K	24	0.03616	-0.09527	0.09742	0.07983	0.80683	0.10019	-0.01069	-0.06600
VALUING	25	0.19582	-0.04296	0.11936	0.21237	0.38542	0.57329	-0.02537	0.09538
RECEIVNG	26	0.22929	0.20196	0.03327	0.21099	0.04540	0.67013	0.00964	0.12975
EFFSOLN	27	0.18917	0.01875	0.09171	0.06397	-0.05857	0.56880	0.09330	-0.38913
INEFFSOL	28	0.29233	-0.05693	0.17960	0.14236	0.09986	0.37265	-0.57074	-0.29083
EMOTADJ	29	0.04795	-0.00117	0.07504	0.07789	0.00851	-0.02742	-0.92897	0.01125

TABLE 7

VARIMAX ROTATION

INDIVIDUAL STUDENT DATA *** GRADE 8 *** <<1974>>

ROTATION OF FIRST 8 FACTORS

COMMUNALITIES

GSELF	1	C.79325	LAWAUTH	15	0.77295
CONTENV	2	C.59120	RESPINTG	16	0.75699
RELATE	3	C.70286	PHEALTH	17	0.57641
SCHLIMAG	4	C.61943	SAFETY	18	0.65007
RACEG2	5	C.61111	DRUGS	19	0.68555
RELIG	6	C.54738	VISLART	20	0.71151
SES	7	C.65130	PERFMART	21	0.75158
INTELL	8	C.64378	SCIENCE	22	0.76769
HANDCP	9	0.67431	WRITING	23	0.70112
GOAL-3V	10	C.78570	GOAL-8K	24	0.69173
GOAL-3M	11	0.73730	VALUING	25	0.59178
LEARN	12	C.70538	RECEIVNG	26	0.60679
SCHOOL	13	C.68840	EFFSOLN	27	C.53495
WELFDIG	14	C.73912	INEFFSOL	28	0.69464
			EMOTADJ	29	0.87793

TABLE 8

SUM OF SQUARED ROTATED FACTOR LOADINGS

	SUM FOR EACH COLUMN	PERCENT OF TRACE
FACTOR 1	3.2255	11.12
FACTOR 2	2.9777	10.27
FACTOR 3	2.4540	8.46
FACTOR 4	2.9833	10.29
FACTOR 5	2.6424	9.11
FACTOR 6	2.9307	10.11
FACTOR 7	1.3167	4.54
FACTOR 8	1.3320	4.59

The Pennsylvania Education Quality Assessment Inventory's efficiency in generating an accurate profile of studentbody needs hinges on the ability of people to communicate with people through the medium of paper-and-pencil tests. Evidence supporting this notion has been obtained through a long series of studies conducted by Department of Education personnel with the help of administrators, teachers and students in over 40 per cent of Pennsylvania's local school districts. Findings support generalizations that:

- Students can read and understand the questions in the battery
- Students tend to answer the questions in such a way as to reflect their true feelings
- Students answer similar items in a consistent manner
- Students tend to answer items in similar ways across time
- Student classroom behaviors are mirrored by student test scores
- Students generally feel the tests are worthwhile and the vast majority take the tests seriously

SECTION FOUR

Target Groups for Program Focus

ORGANIZING INFORMATION TO IDENTIFY STUDENT TARGET GROUPS

Ideally, when preparing to initiate a program to facilitate student progress in any goal area, one should be able to identify students most likely to benefit from that program. However, information available to schools participating in Pennsylvania's Educational Quality Assessment Program does not contain data on individual students. Consequently, it is impossible for school personnel to identify by name the members of the target group toward whom a program might be focused.

Even though individual profiles are unavailable, it is possible to organize data in ways that help identify general student groups that demonstrate needs in a given goal area. This is done by summarizing data for various subgroups of students formed from selected student characteristics. The characteristics defining the subgroups are achievement level, sex and father's occupational status.

Student ability is categorized into three levels on the basis of the composite math-verbal achievement score. Students scoring below the 30th percentile are defined as the low ability group. Students scoring between the 30th and 70th percentile are placed in the middle ability group. Those exceeding the 70th percentile are defined as the high ability group.

Students are also assigned to three groups on the basis of their father's or legal guardian's reported occupation. These occupation categories are labeled for convenience as semiskilled, skilled and professional. These categories are abstractions based upon the average educational requirements necessary to obtain the job and the average amount of compensation for the particular occupations. It is recognized here that there are exceptions in any or all of these categories. The semiskilled occupational category includes hospital attendant laborer, operator of industrial equipment, packer, wrapper, miner, quarry worker, painter, roofer, paper hanger, carpet layer, truck driver, taxi driver, service station attendant, watchman, barber, waiter, cook, farmer and carpenter.

The skilled occupational category included cabinetmaker, dental technician, nurse, librarian, foreman, toolmaker, machinist, electrician, plumber, bricklayer, stonemason, heavy equipment operator, mail carrier, telephone operator, printer, decorator, policeman, firefighter, repairman, butcher, mechanic, tailor, forester, secretary, clerk, office worker, salesperson, grocer and minister.

The professional occupational category includes dentist, doctor, veterinarian, architect, pilot, teacher, school administrator, editor, farm agent, stockbroker, insurance agent, real estate agent, personnel manager, bank official, lawyer, judge, engineer, social scientist and natural scientist.

Eighteen groups are formed by taking all possible combinations of the three student characteristics. The proportion of students who responded favorably to more than one-half of the items comprising each scale are presented in Table 9.

TABLE 9

PER CENT OF STUDENTS SHOWING POSITIVE ATTITUDE: BY GOAL AREA

TYPE OF STUDENTS		
Low ability	Semiskilled fathers	Males
Low ability	Semiskilled fathers	Females
Low ability	Skilled fathers	Males
Low ability	Skilled fathers	Females
Low ability	Professional fathers	Males
Low ability	Professional fathers	Females
Middle ability	Semiskilled fathers	Males
Middle ability	Semiskilled fathers	Females
Middle ability	Skilled fathers	Males
Middle ability	Skilled fathers	Females
Middle ability	Professional fathers	Males
Middle ability	Professional fathers	Females
High ability	Semiskilled fathers	Males
High ability	Semiskilled fathers	Females
High ability	Skilled fathers	Males
High ability	Professional fathers	Males
High ability	Professional fathers	Females
Average per cent showing positive attitudes		

Clearly, in today's world, women are playing an increasingly important role in defining the occupational level of the family. However, data were unavailable to reflect this trend. Therefore, we are forced to use the father's occupational level as a proxy for the socioeconomic conditions of the home.

TABLE 9 (con't.)
PER CENT OF STUDENTS SHOWING POSITIVE ATTITUDE: BY GOAL AREA

GOAL NUMBER	I	II	IV	V	VI	VII-A	VII-P	IX	X
	70	60	52	11	23	61	28	28	70
	66	68	62	19	28	63	17	38	72
	69	64	55	10	20	68	30	31	64
	70	68	65	20	35	70	20	41	74
	69	72	61	18	29	70	30	33	68
	77	68	65	27	33	63	22	43	75
	84	66	61	17	33	60	17	32	76
	78	83	68	34	45	64	13	48	74
	82	67	62	26	38	56	13	41	75
	79	79	66	39	47	63	12	46	79
	86	70	57	20	35	57	15	42	77
	75	81	75	36	41	67	17	54	79
	82	68	73	28	55	53	15	43	79
	86	89	85	54	59	64	19	62	82
	88	71	72	24	50	62	12	52	77
	83	81	79	42	51	66	13	55	79
	89	72	79	25	52	62	16	58	79
	91	93	83	45	53	73	25	72	84
	79%	72%	67%	26%	40%	63%	18%	45%	75%

Note: Student percentages based on random sample of 3,459 8th grade students.

Acknowledgments

Illustrations were composed by Nolan Russell, Division of Educational Quality Assessment, Jan Shaull and Tom Ross, Division of Publications, and Darlene Dalton, a student at Northern High School, Northern York County School District.